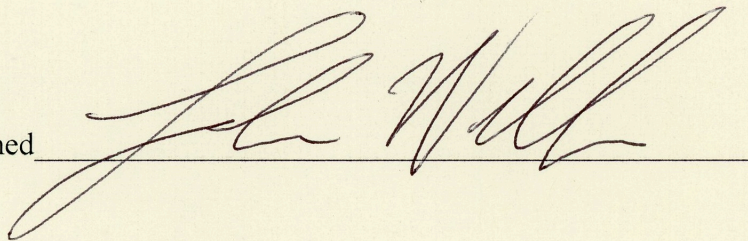


I give permission for public access to my Honors paper and for any copying or digitization to be done at the discretion of the College Archivist and/or the College Librarian.

Signed

A handwritten signature in cursive script, appearing to read "Landon Webber", written over a horizontal line.

Landon Webber

Date

05/21/2014

Examining the Effectiveness and Implications of Teacher Evaluation Policy  
in Memphis and the State of Tennessee

Landon Roch Webber

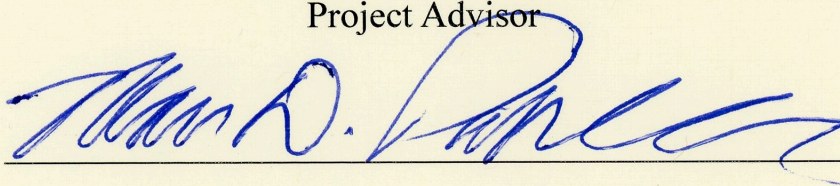
Department of Political Science  
Rhodes College  
Memphis, Tennessee

2014

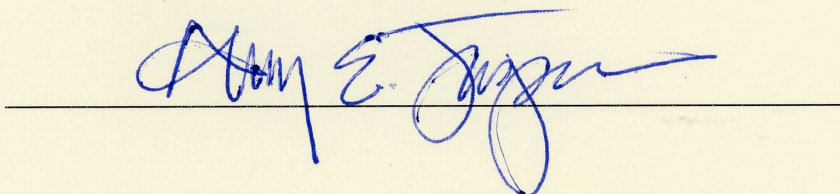
Submitted in partial fulfillment of the requirements for the  
Bachelor of Arts degree with Honors in Political Science

This Honors paper by Landon Webber has been read  
and approved for Honors in Political Science.

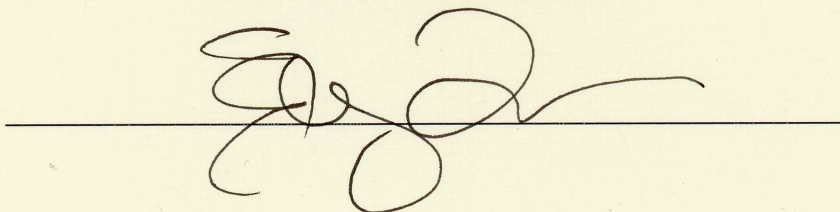
Dr. Marcus Pohlmann  
Project Advisor



Dr. Amy Jaspersen  
Second Reader and Department Chair



Dr. Elizabeth Thomas  
Extra-Departmental Reader



## ACKNOWLEDGEMENTS

This project would be sorely incomplete without the support and participation of a network of faculty members, community leaders, school district officials, teachers, principals, and fellow students who made this work possible. I would be remiss if I did not begin my report with acknowledgement of the contributions of the large number of individuals who assisted me in completing this report. I would first like to express my sincere appreciation advisor, Professor Marcus Pohlmann, for his constant guidance throughout my work on this project. Professor Pohlmann sat through countless meetings, endured and answered endless questions, provided much-needed direction and focus to my research and offered me essential contacts in the community, school districts and state in order to allow me to make a substantial contribution to education policy research through my work. I would also like to thank Professors Amy Jasperson and Elizabeth Thomas for their encouragement and constructive feedback, which always came in the form of concrete, understandable suggestions of how to improve. Additionally, I would like to thank the many other faculty members who have greatly influenced the outcome of this research project. Professor Lindsay Flynn, my research mentor, sparked my initial interest in education policy research and welcomed my questions and curiosity, helping to hone and perfect my research questions. Professor Timothy Huebner was the first to suggest that I consider pursuing Honors Research and offered advice on how to begin the process for proposing my topic. Professor Daniel Cullen offered his support and encouragement as my academic advisor. Professor Renee Johnson encouraged me, showing eager interest in my preliminary research,

encouraging me to present findings at the Midwest Political Science Association 2014 Annual Conference and always brought a thoughtful, challenging and analytical critique to my project.

My interest in and passion for studying issues of inequity in American public education would not have been fostered or developed were it not for the support and encouragement of the Bonner Foundation and the Bonner Center for Faith and Service at Rhodes College. Bonner Center staff and students have consistently encouraged me in my work and have often allowed me to see tangibly the importance of understanding and thoughtfully confronting the challenges facing equitable public education, particularly in Memphis. I would like to thank Shannon Hoffman and Chaplain Walt Tennyson for their support of my work and their ability to help me make meaningful connections between my classroom lessons, my career interests, my work in the Memphis community and my research.

I would like to express my gratitude to all the staff, students and faculty of the 2013 Rhodes Institute for Regional Studies. My selection and my inclusion in this thoughtful, dedicated and creative cohort of students, through which I was able to serve as a summer Research Associate for Shelby County Schools, offered the opportunity for me to begin thinking earnestly about my project. My gratitude goes out to Professor Milton Moreland for his leadership of one of Rhodes' most engaging and exceptional student programs. I would like to thank my advisor for my Rhodes Institute project, Professor Marcus Pohlmann, for setting up the research internship at Shelby County Schools and for guiding me throughout the research for and writing of that report, which served as the early stages of my Honors project. I am sincerely grateful for the support and encouragement of Shelby County Schools staff over the course of

this project as well, most notably that of Dr. Laura Link, Mr. Eric Linsy, Ms. Jessica Lotz, Ms. Jody Lanza, Ms. Kemmashela Smith and Ms. Valarie McCoy.

In assisting with this project and making my analysis of teacher evaluation policy in Memphis, Shelby County and the State of Tennessee possible, I would like to thank the staffs of the Offices of Performance Management and Planning and Accountability at Shelby County Schools. To Ms. Rachel Lebo and Ms. Emily Marquart, thank you for being willing to explain the intricate details of teacher evaluations to a newcomer. To Ms. Jessica Lotz, Mr. John Hester, Dr. Tracey Wilson, and Dr. Brant Riedel, thank you for providing me with the data and permission I needed to proceed with my research. Ms. Lotz was especially helpful in her assistance with my project and with advising me on who to interview, providing teacher contact information and giving me access to past surveys and teacher evaluation research. Last, but not least, I would like to thank fellow students William McGriff and Corey Lipschutz for their helpful support and assistance, even with such short notice, helping me to conduct an analysis of the teacher evaluation and survey data provided by the school district. Without the help and assistance of all of these individuals and institutions, I would most certainly not have been able to accomplish this work.

## CONTENTS

Signature Page	ii
Acknowledgements	iii
Contents	vi
List of Tables and Graphs	ix
Abstract	x
Introduction	1
Background and Context: Memphis and Shelby County Schools	5
Teacher Evaluations in Memphis and the State of Tennessee: A Policy History	31
Review of Relevant Literature	
<i>The History of Performance Evaluations</i>	44
<i>The Introduction of Performance Evaluations to Government     Institutions</i>	48
<i>The Introduction of the Concept of Performance Measurement     in a Public Education Context</i>	54
<i>The Introduction of Teacher Performance Evaluations in     the United States</i>	59
<i>The Purposes for Implementing Performance Evaluations</i>	65
<i>The Importance of Implementation, Communication and     Creating a Culture of Feedback</i>	73
<i>Are Performance Evaluations Useful and Under What     Circumstances?</i>	79

<i>The Issue of Accuracy in Performance Evaluations</i>	85
<i>The Issue of Employee Morale and Motivation</i>	88
<i>Issues Involving Communication and Implementation in Performance Evaluation Systems</i>	91
<i>The Manager-Employee Relationship and its Effects on Performance Evaluations</i>	96
<i>The Cost of Maintaining and Implementing a Performance Evaluation System</i>	98
<i>The Significance of Teacher Evaluation Policy to the Education Reform Debate</i>	101
<i>Effects of Changes to Teacher Licensure and Certification on Evaluation Policy</i>	104
<i>What is Teacher Quality?</i>	107
<i>The Use of Value-Added Modeling to Assess Teacher Quality</i>	114
Methodology and Approach	
<i>Introduction</i>	117
<i>Teacher Focus Groups</i>	118
<i>Data Analysis</i>	122
<i>Research Questions</i>	123
<i>Hypotheses</i>	124
<i>Approaching the Issues</i>	129
Discussion of Results	
<i>Student Achievement in Memphis and Shelby County Since the Adoption of the Teacher Effectiveness Initiative in 2010</i>	132



<i>Teacher Perceptions of the Teacher Evaluation System and Its Effectiveness in Memphis, Shelby County and the State of Tennessee</i>	138
<i>Reliability of the “Multiple Measures” of Teacher Evaluations</i>	155
<i>Side Effects of Teacher Evaluation Policy: Teacher Morale, Motivation and Commitment to Teaching</i>	166
Conclusion	176
Tables and Graphs	180
References	187

## LIST OF TABLES AND GRAPHS

Table 1. <i>2009-2013 State Teacher Evaluation Legislation</i>	180
Table 2. <i>TCAP Criterion-Referenced Academic Achievement: Math, Grades 3-8</i>	181
Table 3. <i>TCAP Criterion-Referenced Academic Achievement: Reading/Language Arts, Grades 3-8</i>	182
Table 4. <i>2013 Teacher Evaluation Summative Scores for Memphis City Schools</i>	183
Table 5. <i>2012 Teacher Evaluation Summative Scores for Memphis City Schools</i>	184
Table 6. <i>2013-2014 Partial Teacher Classroom Observation Scores (Rated by Principals)</i>	185
Table 7. <i>Percentage of Teachers Rated as “4” or “5” on Classroom Observation, TVAAS and Student Achievement Data Scores in 2011-2012 and in 2012-2013</i>	186

## ABSTRACT

### Examining the Effectiveness and Implications of Teacher Evaluation Policy in Memphis and the State of Tennessee

by

Landon Roch Webber

Since the 1980's, a growing network of those invested in the debate over reforming public education in the United States has consistently pushed for the development and implementation of teacher evaluation systems based heavily on student standardized testing data. Twenty-three states now require or recommend that student achievement data comprise half of a teacher's evaluation score. As a result of this "sea change" in policy, researchers have come to identify data-based teacher evaluations as essential components of the "ideology of school reform." By 2010, Memphis was positioned to serve as a model of rigorous teacher evaluation design for the rest of the country.

A partnership with the Gates Foundation, inclusion in the Measures of Effective Teaching study and support from state officials completing Tennessee's Race to the Top proposal produced a system which scores teachers largely on the basis of student achievement data

and classroom observations by principals. In its third year, the evaluation framework has garnered national attention and now applies countywide as a result of the recent merger with Shelby County Schools. The system, however, is highly controversial with teachers, who see test scores as unstable measures of their contribution to student learning. Examining district evaluation data and utilizing surveys, interviews and focus group discussions with teachers, this paper offers a preliminary assessment of the effectiveness of this evaluation framework and its implications for teaching and learning in inner city schools, both in Memphis and more generally.

## Introduction

The modern-day education reform movement undoubtedly aims to improve the quality of instruction students receive in public school classrooms across the country, regardless of location, race, gender, background, income level or socioeconomic status. While all of these students are guaranteed the right to attend public schools in their districts, free of charge, the education reform movement has pressed beyond this relatively straightforward guarantee to question the effectiveness and equity of the educational opportunities available to students on a day-to-day basis. We are just past the thirty-first anniversary of the publication of *A Nation at Risk*, the monumental government commission report which shined a light on the shortcomings of the American educational system in *actually* educating its students and preparing them to be both effective, competitive members of the workforce as well as responsible, compassionate and thoughtful citizens. The report highlighted what it decried as the “rising tide of mediocrity” in American public schools, identifying lagging student achievement, measured primarily in the form of trends in SAT scores, as one of the symptoms of the problem (National Commission on Excellence in Education 1983). The report placed a spotlight on American public education as the object of public policy activity and, for this reason, remains one of the few government commission reports mentioned in routine discussions of 20th century American history. As a result of the report, an energy and focus were brought to the discussion of how to reform and improve the nation’s education system which had not been present before. And, since the 1980’s, a steady stream of local, state and national policymakers, parents, teachers and teachers’ unions, researchers and a host of other interest groups and interested parties have kept the conversation going. It is a policymaking arena that has seen both dynamism and longevity, having remained a

consistent focus of public policymaking since the 1980's. The outputs of this policymaking can be measured in phases. This paper seeks to address one of the education reform movement's most recent phases- the push to assess teacher effectiveness and performance. The early beginnings of this policy trend began in the late 1980's and 1990's, but since roughly 2009, policymakers, private consulting and research firms and lobbying groups have targeted and pushed for the development of teacher evaluations based on "multiple measures" of which two key components are principal classroom observations of teachers, usually annually if not more frequently, and student standardized testing data.

Teacher evaluation is certainly one of the most important and most long-lived focal points of the education reform movement. But why is it the object of such emphasis? The answer lies in the fact that teacher evaluations are at the heart of the education reform movement's focus- classroom and instructional quality. The push for ensuring provision of *quality* instruction and learning opportunities to all American students as opposed to merely providing free schools was first described by Robert Hutchins, educational philosopher and President of the University of Chicago in the 1930's. He stated: "Perhaps the greatest idea that America has given to the world...is the idea of education for all. The world is entitled to know whether this idea means that everybody can be educated, or only that everybody must go to school" (Ravitch and Viteritti 26). Understanding Hutchins' distinction, the distinction between *schooling* and *education* is crucial to understanding many of the recent surges in education policymaking. If policymakers can find a way to assess the performance of teachers in the classroom and use these assessments to improve the quality of the instruction students receive, then, as the argument goes, they will be able to truly improve American public education by hitting at the heart of the issue- poor

instructional and educational quality. This paper assesses the results and implications of a teacher evaluation system born out of this policymaking wave and currently in its third year of implementation. In 2010, Memphis City Schools began the development of its Teacher Effectiveness Initiative. Funded by the Gates Foundation and driven by local administrators who saw the potential to put Memphis on the map of education reform, the district developed a performance evaluation system for teachers, strict, according to national standards, and based in large part on student standardized testing data and principals' observations of teachers' classroom performance against a standardized rubric known as the Teacher Effectiveness Measure (TEM). These innovations coincided with the State of Tennessee's application to the U.S. Department of Education's nationwide Race to the Top competition. In 2011, Tennessee and Delaware were the first two states to win funding. Since then, Tennessee and Memphis in particular have been hailed as national exemplars in the field of teacher assessments. In this way, they serve as the ideal subjects of a teacher evaluation policy case study. They also offer a unique opportunity to assess a teacher evaluation policy crafted by or at least heavily influenced by some of the most prominent leaders of the education reform movement itself.

What is provided below is an analysis of the history, implementation and effects of teacher evaluation policy in Memphis, Shelby County and the State of Tennessee. Like any analysis of public education policy, the report is filled with many qualifications. What is perhaps made most evident by this study is that teaching and learning is a complex process with many contributing factors, or inputs, and many different ways to measure, assess or understand its outcomes. There are significant limitations both to the conclusions drawn below and also to the evidence and information used to reach these conclusions. Regardless of this uncertainty in our

inability to measure key aspects of the educational process, however, we know that studying education with the intent of improving the teaching and learning process is a worthwhile endeavor. The benefits are real and can improve the lives of actual teachers and actual students in actual classrooms. This has the potential, as we know, to make an actual difference. Despite its limitations and shortcomings, this report offers an analysis of one of the most significant and far-reaching attempts to measure and improve educational quality in public schools to date.

As the report also notes, this policy and its implementation are not fully understood without a brief introduction to the unique context of this adoption and implementation: Memphis, Shelby County and Tennessee, and the recent struggle over the school district merger and de-merger. To this end, a short background of education in Memphis, particularly focusing on the largest school system merger in American history, which began in 2010 and culminated with the unification of Memphis City Schools and Shelby County Schools, the smaller, suburban school district, is provided. This merger had significant effects on the work of the Teacher Effectiveness Initiative. Next, the paper offers a brief history of how the teacher evaluation policies currently in place were developed and what existed before their enactment.

Particular insight is offered into the relationship between the school district in Memphis and the State of Tennessee in crafting these policies, into the involvement of the Gates Foundation in sponsoring and overseeing the cutting edge research that occurred involving Memphis schools, teachers and students and into the implementation of the complex and highly-detailed system that has been put in place since 2011. A review of relevant literature is conducted which examines how teacher evaluations came to be an important element of education reform. The paper concludes with a discussion of the contributions this study makes to the policy



analysis of teacher evaluations. Results from this research and an overview of surveys of teachers that have been conducted by the district since the evaluation program was implemented are discussed. This study offers a more in-depth examination of the complexities and difficulties facing a large urban school district attempting to implement such a set of policies. Because it is only with well-considered and thoughtful analyses of policy programs and their successes, shortcomings and oversights that policymakers and practitioners can ever hope to improve upon their work and move closer toward achieving their goal of improving educational quality in the public schools, I offer this report in the attempt to aid the ongoing effort to improve the quality of teaching and learning in the public schools.

### **Background and Context: Memphis and Shelby County Schools**

From 1826 until 1848, all education in Memphis occurred in the homes of “well-to-do citizens,” who had been chartered by the state to offer education privately (Memphis City Schools 2007). In 1826, the first of these citizens was Mr. Underwood from Alabama who taught at the first “school” at his home in Memphis (Memphis City Schools 2007). “Other schools followed taught by businessmen and clergymen in their businesses, churches and homes” (Memphis City Schools 2007). This system obviously did not cater to “lower- to middle-class white children” and, in this antebellum period, schools were “non-existent for black children” (Memphis City Schools 2007). When public schools were first offered in Memphis, however, this all changed. While the state passed its first legislation offering public schools to students in 1836, the district did not open its “first free school,” located at Third and Overton

until 1848 (Pohlmann 2008, Memphis City Schools 2007). Eventually, with the help of J.W.A. Pettit, Memphis' Board of Aldermen was encouraged to begin funding the new schools with the use of city treasury funds (Memphis City Schools 2007). The district's first year budget amounted to \$20,000 (Memphis City Schools 2007). In 1852, the first "city school tax rate" was imposed on residents, followed by the first "county school tax rate" in 1854 (Memphis City Schools 2007). By this point, the district boasted 13 schools (Memphis City Schools 2007).

As Memphis City Schools stated in a brief history of the district which appeared on its website until its dissolution in 2013, "The creation of these tax rates allowed children who formerly could not afford school tuition to receive a public education, but it would still be another 20 years before public education was extended to black citizens" (Memphis City Schools 2007). Memphis' extension of the right of public education to black students did not start until 1868, at Clay Street School, the "first recorded black public school" founded in 1873 (Memphis City Schools 2007). The first class to graduate from a black public high school graduated in 1891 (Memphis City Schools 2007). By 1890, "40 percent of the school-age population of Memphis [was] enrolled in public school" (Memphis City Schools 2007). By 1910, the number had risen to "80 percent" for white school-age students and "60 percent of black children aged 6 to 14 [were] enrolled in school" (Memphis City Schools 2007). In the 1800's, Memphis City Schools petitioned to "become its own school district" and was granted "special school district status by the state" (Canon 2011). This meant that the district was still part of Shelby County, but was separate from the administration and funding structure of the county's school district, Shelby County Schools (Canon 2011). During this period, within Memphis' own special school district, schooling remained segregated. This ended on January 24, 1973 when court-ordered busing

began in Memphis (Memphis City Schools 2007). The integration process caused “a great deal of controversy,” particularly among white parents (Memphis City Schools 2007). In January alone, over 7,000 students withdrew from the public schools, enrolling in temporary schools set up by Citizens Against Busing or in Memphis-area private schools (Memphis City Schools 2007). Since that time, through a series of annexations, Memphis City Schools grew into a large, urban school district, the 21st largest in the nation, serving over 100,000 students and 187 schools (*The Commercial Appeal* 2011). This was the status of the district on the eve of its merger with Shelby County Schools.

On November 22, 2010, Memphis City Schools board members Martavius Jones and Tomeka Hart put forth a resolution aimed at surrendering the school system’s charter and placing the control over the district’s 187 schools and nearly 105,000 students in the hands of the much smaller, suburban Shelby County Schools system (Roberts 2010, *The Commercial Appeal* 2011). The move, made by a former president of the Memphis City Schools board, seemed at first an odd one for a system “struggling to raise student achievement and fighting with the city council for money” (Roberts 2010). Then president of the board, Freda Williams, saw the move as a distraction from the more pressing work of the system, noting that, “Our attention is divided, unfortunately, but we cannot stop working toward high achievement for all of our students” (Roberts 2010). Some board members took offense to the resolution. Board member Kenneth Whalum asserted that those who supported the move “should resign immediately, and clear the way for the appointment of school board members who are ready to do what our charter empowers us to do” (Roberts 2010). For proponents of the resolution, however, the initiative was not without cause. Shelby County Schools, the school system which served all students in Shelby

County outside the city limits of Memphis, had recently attempted to redefine itself, not as a countywide school district from which Memphis was excepted, but rather as a special school district in its own right as Memphis City Schools was. Under the system set up to fund the two districts, “all Shelby County residents, including Memphians, [paid] taxes to the county” (Canon 2011). The county commission was then tasked with distributing “funds between the school systems based on the number of students who attend” (Canon 2011). On top of this, Memphis City Schools “then [provided] additional funding beyond this to its schools because of its special status” (Canon 2011). It was because of Shelby County Schools’ thinking about redefining itself as a special district which made school administrators in Memphis nervous about losing the possibility to be included in receiving revenue from the tax base outside the city’s corporate limits which caused board members Jones and Hart to consider dissolving Memphis’ own special school district status, thereby preventing the county school system from forming its own district and consolidating the two districts and the source of funding.

As Jane Roberts, reporter for *The Commercial Appeal*, wrote, “The city school board’s charter debate is a reaction to a plan by Shelby County Schools to seek designation as a special school district” (2010). Shelby County Schools’ efforts to declare itself a special district posed financial problems for Memphis City Schools, some argued, since, as a result of the move, taxes paid by residents of parts of the county lying outside the city limits would only go to support suburban schools and could no longer be used to fund both suburban and city schools as they had been since the beginning of the use of separate city and county school systems (Roberts 2010). Because of the greater wealth and high property values outside the city, this represented a significant loss of revenue for Memphis City Schools (Roberts 2010). A 2008 University of

Memphis study noted that, should Shelby County Schools become a legislatively-approved special district, “city taxes for education subsequently would have to rise 20-23 cents per \$100 of assessed valuation for city property owners to cover costs of educating the 70 percent of children in the country who live within the city boundaries” (Roberts 2010). The proposal was a highly controversial subject both in and outside the city limits. Memphis residents argued that the board’s debate over whether or not to surrender its charter and be absorbed into the Shelby County Schools system was a distraction to the larger, more pressing challenges for public education facing Memphis. Suburban residents, fearful of what a wholesale system merger might mean for their schools, protested the dissolution of the charter in large numbers (Roberts 2010).

On December 20, 2010, after a six hour meeting filled with “squabbles over parliamentary procedure and biblical references laden with more emotion, pathos and angst,” the Memphis City Schools board voted five to four to surrender its charter should the proposal be confirmed by the city’s voters in a referendum (Roberts 2010). Immediately, leaders of Shelby County Schools argued that the decision should be reversed, pleaded with the Memphis City Schools board to rescind its decision and stated that the “‘shotgun marriage’ or forced consolidation” would be disadvantageous for the students of both systems (Silence and McMillin 2011). Chairman of the Shelby County Schools board David Pickler promised that “the system will consider every legislative and legal option to ‘protect the interests of Shelby County school students and taxpayers’” (Silence and McMillin 2011). In January, Shelby County school board members gave administrators “the green light...to take any legal action they deem necessary to block consolidation with city schools” (Silence and McMillin 2011). Nevertheless, the resolution to offer the issue up to voters in a referendum went forward. In February 2011, the Memphis City

Council approved the charter surrender and continued to support allowing voters to decide the issue in a city-wide election (Locker and Maki 2011). In anticipation of the merger, the Shelby County Commission voted to expand the size of the Shelby County Schools board from seven to twenty-five members in order to make the board more representative of its newly-formed district, over the great protestation of Shelby County Schools officials (Connolly 2011). On March 8, roughly 67 percent of voters approved the surrender of the charter and authorized the beginning of the process of merging the two districts (McMillin and Roberts 2011). Despite the large margin of voters who supported the merger, only roughly 17 percent of Memphis voters turned out to vote during the referendum (McMillin and Roberts 2011).

As the Shelby County Commission began thinking about how to organize the structure of the new unified district and how to oversee what would become a long transition period, Shelby County Schools took the matter to federal court, arguing in its lawsuit that U.S. District Court Judge Mays “to either strike down the MCS charter surrender or to make clear when a transfer of MCS to the county schools would take effect” (Buser 2011). Proponents of the merger, including Memphis City Schools board member Martavius Jones, who had first introduced the resolution to dissolve the city school system’s charter, were astonished by Shelby County Schools’ persistence in attempting to avoid and contest the merger. Jones stated: “The outrageousness of all this is that Memphis is part of Shelby County. We’re using Memphis tax dollars to sue Memphis out of its right and the obligation Shelby County Schools has to educate all children in Shelby County” (Buser 2011). A law which had been “recently pushed through” the Tennessee State Legislature during the spring legislative session by Senator Mark Norris of Collierville delayed any merger until 2013 and required the oversight of a 21-member Transition Planning

Commission or TPC (Buser 2011). In August 2011, Judge Samuel Mays ruled that the surrender of the Memphis City Schools charter was legal and that the merger effort would move forward, canceling the attempts of Shelby County Schools to put an end to the situation through litigation (McMillin 2011). Judge Mays also ruled, however, that the Norris-Todd law delaying the merger and outlining its process was legal and that the procedures outlined in the law would be followed (McMillin 2011).

Mays ruled that the consolidation must be complete by the 2013-2014 school year (McMillin 2011). Judge Mays also ordered the parties in the lawsuit to move swiftly in coming up with a plan for how the newly-unified school board might be composed of members representative of the new communities serviced by the district both in and outside the Memphis city limits (McMillin 2011). Soon afterward, county and city school officials produced a plan to name a 23-member unified school board to oversee and implement the proposals of the Transition Planning Commission (McMillin 2011). Ultimately, the final merger deal, developed over months of mediation sessions between administrators of the two school systems led by Judge Mays, was approved by both school boards unanimously in August 2011 (McMillin and Roberts 2011). What was once a highly controversial and acrimonious issue had now been made, at least for the most part, agreeable to the administrators of each school system. Even the chairman of the Shelby County Schools' board David Pickler who was a longtime and dedicated opponent of the merger, seemed mostly satisfied with the deal. He stated that the agreement over how the merger would proceed with Memphis City Schools was "the first clarion call to this community that we are looking for a brighter future for Memphis and Shelby County for every child" (McMillin and Roberts 2011). Under the guidance of the Transition Planning

Commission, Judge Mays and the unified school board, the two school districts began a consolidation process which would last almost two years.

By the summer of 2012, the Transition Planning Commission had produced their final set of recommendations for how the merger ought to be approached. As *The Commercial Appeal* reporter Michael Kelley noted, “After spending an estimated 400 hours in 150 meetings with 100 members of the Memphis City and Shelby County Schools staffs, interacting with about 14,000 in community meetings and producing more than 10,000 pages of documents, Transition Planning Commission members believe they have an acceptable plan for merging the districts” (Kelly 2012). In order to begin to move toward the operation of a nearly 147,000 student system, the Transition Planning Commission proposed, among other things, a “multiple achievement paths model” which would seek to give educational leaders, schools and regions more autonomy in making school decisions (Kelley 2012). In addition, the proposal attempted to “improve services and lower costs” by suggesting that “the number of employees required to staff the central office” be reduced, “class sizes” be reset to accommodate the needs and meet the standards of those students coming from either city or county schools and also that the “salaries and benefits of the districts’ employees” be harmonized (Kelley 2012). All in all, the plans proposed by the Transition Planning Commission were intended to make way for a system which would “serve every public schoolchild in Shelby County” (Kelley 2012). Throughout this period, the two school districts continued to operate as though they were still separate school systems. Officials from the two school systems began to meet more frequently to prepare for and begin to execute the Transition Planning Commission’s initial recommendations. As late as February 2013, however, Judge Mays, “concerned about the pace of the schools merger,...ordered all



parties involved to make recommendations about appointing a special master who could have broad authority to insure operations of Memphis and suburban are combined” in time for the July 1, 2013 merger deadline (Bailey 2013). Judge Mays wanted to give the special master authority to create and enforce hard deadlines and to propose courses of action which he or she could take directly to the judge without consulting the authorities of either of the merging school systems (Bailey 2013). By March, Judge Mays had assigned former Memphis city chief administrative officer Rick Masson to the role of “special master” overseeing the day-to-day operations of the schools merger (McMillin 2013).

As the merger deadline moved closer, however, the school systems began to accomplish more and more of the necessary tasks and, on July 1, Memphis City Schools, a district which had existed since 1848, was dissolved and the newly-unified Shelby County Schools assumed educational responsibility for the roughly 150,000 public school students in Shelby County (Kelley 2013). Currently, the combined district, now serving students from both the city and the county’s suburbs, is the 14th-largest school district in the nation (Dillon 2011). In the eyes of the nation, the school merger loomed large in conversations over education reform and restructuring. Sam Dillon, writing for the *New York Times*, called the merger of the “overwhelmingly black Memphis school district” with the “majority-white Shelby County schools...the largest school district consolidation in American history” (Dillon 2011). Perhaps what attracted the most national attention to the merger were the characteristics of the school district which had already been attracting non-profits, education reform institutes and think tanks, advocacy organizations and national charter school networks. Even in the years preceding the merger, the natural characteristics of this urban school district were what had turned Memphis into a hub of national

educational reform and ingenuity- its sheer size, its large number of minority and economically disadvantaged students and its seemingly chronic battle with poor student performance. As Michael Kelley summarized of Memphis City Schools shortly before the two districts merged, “It is one of the largest, poorest and most troubled districts in the country” (Kelley 2013). This represents why so many of the nation’s top education reformers have descended on Memphis in recent years. They want to see if their ideas, programs and institutions will be able to transform and reinvigorate public education in a district which serves what is possibly the country’s most vulnerable, at-risk student population. In response, Memphis has frequently been willing and open to implement and try some of the most radical programs of reform in the country, not least of which is the new Teacher Effectiveness Measure developed in concert with the Bill and Melinda Gates Foundation. All of these factors create an environment which fosters and encourages educational innovation and attracts national ideas, proposals and grants. It was in this already reform-oriented environment that Memphis and Shelby County began their historic merger.

Initially, some were skeptical of the ability of the two districts to ever come to terms with and unify their distinct approaches to almost every administrative and maintenance aspect of running a large, urban, consolidated school district. As Dillon noted, the merger posed “huge logistical challenges” (Dillon 2011). He stated: “Memphis teachers are unionized, Shelby County’s are not; the county owns its yellow buses, the city relies on a contractor; and the two districts use different textbooks and different systems to evaluate teachers” (Dillon 2011). To be sure, the two districts were very different, perhaps most notably in the demographics of the student bodies they each served. In 2010, Memphis City Schools served just over 103,000

students in 187 schools, while Shelby County Schools served over 47,000 students in 53 schools (*The Commercial Appeal* 2011). In 2010, 85% of the student population of Memphis City Schools was African American and 87.2% of its students were also economically disadvantaged (*The Commercial Appeal* 2011). In comparison, roughly 38% of Shelby County Schools' students were African American, while 37% of the system's students were economically disadvantaged (*The Commercial Appeal* 2011). The nature of the overall populations in the two school districts, and not just of their student populations, was markedly different. The median annual household income for families living within the Memphis City Schools district was \$35,535 and almost 40% of the district's residents under the age of 18 were living in poverty (*The Commercial Appeal* 2011). The median annual household income for families living within the reaches of the Shelby County Schools district was, in contrast, \$83,998 and only 8.2% of the district's residents under the age of 18 were living in poverty (*The Commercial Appeal* 2011).

Perhaps one of the most prominent differences in the school districts lay in the academic achievement of their students. This prompted many to label Shelby County Schools from the beginning of the merger debate as "the more successful" of the two districts in terms of student achievement (Kelley 2013). As the opponents of the merger fought to have the process halted or delayed at every turn until the consolidation of the two districts became inevitable, they often cited being fearful of having the academic strength of their district brought down by having to take the poor performing and often failing city schools under their wing. Joseph Clayton, who served on the board of the "mostly suburban" Shelby County Schools before the talk of the merger began and who was eventually elected to the unified school board which oversaw the transition process, noted that most of the opposition to the merger was motivated by this fear of

losing control of what had been a small, manageable, high-performing school district (Dillon 2011). He explained on the eve of the consolidation: “Today the fear is about the academic decline of the Shelby Schools” (Dillon 2011). By the state’s standards, nearly 90% of the county schools were labeled as being in good academic standing, while only roughly 42% of the city’s schools received the same ranking (*The Commercial Appeal* 2011). In addition, Shelby County Schools had maintained a graduation rate of nearly 92%, while in Memphis, the graduation rate was slightly over 70% (The Commercial Appeal 2011). In 2012, writing for the *Memphis Flyer*, John Branston compared the average number of students deemed proficient or advanced in either reading or math. He noted that, for students of Memphis City Schools, in students in grades 3-8, nearly 30% were proficient or advanced in both math and reading according to test results (Branston 2012). By comparison, Shelby County Schools students in the same grades tested at higher levels, with roughly 57% of students testing as proficient or advanced in math and over 60% testing at the same level in reading (Branston 2012).

For students in high school, the results were not much different. Roughly 33% of students in Memphis tested at acceptable levels in algebra, compared with over 60% in Shelby County (Branston 2012). In Memphis, over 43% of students were deemed proficient or advanced in English, compared with over 70% of Shelby County Schools students (Branston 2012). These drastic difference in student achievement statistics had remained for years, even though Memphis City Schools spent more per pupil and had a slightly lower student-teacher ratio than did Shelby County Schools (*The Commercial Appeal* 2011). Some, including activists such as Kenya Bradshaw with Memphis’ branch of the national educational advocacy group Stand for Children, claimed that both systems suffered from a tendency to deliver inadequate educational services to

students (Dillon 2011). At one of the Transition Planning Commission meetings, representatives from Stand for Children argued that, according to the results from ACT testing, the number of students from Shelby County Schools who were deemed college ready according to the test was not very high (McMillin 2011). According to testing data from the 2009-2010 school year, while only 5% of Memphis City Schools students were assessed as able to proceed to college, only 20% of Shelby County Schools students received the same ranking (McMillin 2011).<sup>1</sup> For activists like Bradshaw, this represented an opportunity for both Memphis and Shelby County to “re-envision” their “educational system” together, sharing ideas about reform and effective teaching and learning and, in the end, improving both systems as a result (McMillin 2011).

These developments brought on by the merging of Memphis City Schools and Shelby County Schools have affected the implementation of the Teacher Effectiveness Measure as it was originally laid out and developed and as it was later applied to the merged district. Writing for the *New York Times* and covering the district merger, Sam Dillon noted that one particular challenge for the new district to consider would be how to evaluate its teachers since, as he

---

<sup>1</sup> It is important to be careful with these numbers. According to the definitions of “college readiness” used for these figures, students are deemed college ready when, on each section of the ACT (English, math and science), they receive the score which indicates proficiency. The scores required to demonstrate proficiency (and thus college readiness) differ for each of these subjects. On the English section, for example, students who receive a score of 18 are deemed to be college ready. On the Mathematics section, however, students who receive a score of 22 are deemed college ready and on the Science section, the required score is 24. The percentages above are of those students in each district who received a proficiency score *on each section of the ACT exam*. With the ACT, however, one score is ultimately produced from the average of the three sections. If students were to meet the ACT’s benchmark score for college readiness on each section, their average cumulative score would be a 21. This cumulative score, however, is higher than the score used by most universities to assess college readiness. At most universities, a cumulative score of 19 is used to draw the line between those students who are ready for college and those who are not. In 2011, as arguments were made by both districts about how each district was preparing their students for college, Shelby County Schools argued, and compellingly so, for the use of the benchmark cumulative score of 19 instead of the section-by-section approach to the ACT’s standards for college readiness. When a cumulative score of 19 is used to determine college readiness, over 62% of Shelby County Schools juniors in 2010 received this score and were deemed college ready. By comparison, Memphis City Schools noted that, overall, a record number of students had scored a 19 or above- 2000 students in total. This showed, however, a much lower percentage of juniors and seniors deemed “college ready” according to ACT standards in Memphis City Schools than in Shelby County Schools.

explained, Memphis City Schools and Shelby County Schools each had “different systems for evaluating teachers” (Dillon 2011). The issue of teacher evaluations, especially in light of the three year long effort to reform Memphis’ system with the Teacher Effectiveness Initiative, supported through the assistance of the Gates foundation grant, was something which had to be considered and assessed. Memphis City Schools and the Gates Foundation both noted, as the merger moved forward, that the Teacher Effectiveness Measure and its continued implementation, as well as the funds provided for its development by the Gates Foundation, would not be effected by the merger, as long as the city continued to prove its commitment to reforming teacher evaluations in order to adhere to the goals and framework already established by the Teacher Effectiveness Initiative. The Gates Foundation publicly declared its commitment to continuing its work with teacher evaluation reform in Memphis, stating: “The Gates Foundation is committed to funding the Teacher Effectiveness Initiative for our children...The foundation has pledged its support as long as effective teaching and improved outcomes for all students remain a priority. The merger vote has done nothing to change our focus in these areas: we remain hard at work on this important reform effort” (Memphis City Schools 2011). Over the spring and summer, the district worked to revise and adjust the Teacher Effectiveness Measure and the Leadership Effectiveness Measures. The Teacher Effectiveness Measure is currently in its third version and, to date, teachers in the newly-unified Shelby County Schools system are evaluated on how well they measure up against a rubric consisting of seven criteria defining effective teaching, four criteria relating to how well the teacher “cultivates a learning environment” and four standards of professionalism for teachers. The current principal evaluation seeks to hold principals accountable for their daily performance on the job as

observed by their direct supervisors, student growth and achievement data and also the quality of the teacher evaluations they conduct.

A consideration of how these developments have shaped education reform and teacher evaluations in the greater Memphis area is important in understanding the district's current approach to teacher evaluations. Administrators at Shelby County Schools have been clear to stress that as the district moves forward, its evaluation programs for teachers and principals are one of its key policy foci and the values behind these systems serve as some of the foundational guidelines the district uses in creating new policies. In the summer of 2013, for example, when the district began the long process of creating a uniform performance evaluations system for all non-instructional staff including nutritionists, cafeteria workers, nurses and maintenance workers as well as cabinet level administrators, it stressed that such a system ought to be clearly, purposefully and consciously linked to the evaluations already conducted for teachers and principals (Shelby County Schools 2013).<sup>2</sup> The original concept of administrators and other non-instructional employees being held to the same standards and being subject to performance evaluations in the same way that teachers and principals are held accountable by the Teacher Effectiveness and the Leadership Effectiveness Measures was originally brought to the forefront of the agenda of the Office of Performance Management during the months prior to and directly following the merger *due to* complaints from teachers who argued that district non-instructional classified and administrative staff were subject to minimal, cursory performance evaluations, while teachers were held to ever more rigorous standards, if such staff were evaluated at all (Shelby County Schools 2013). In the eyes of the district's teachers, this did not seem fair,

---

<sup>2</sup> This is according to a personal interview with an upper-level staff member in the Office of Performance Management at Shelby County Schools prior to the finalization of the school district merger in July 2013.

especially since administrators and non-instructional staff played significant roles in directly affecting student achievement themselves or were essential to supporting the work of the teachers themselves in improving student achievement (Shelby County Schools 2013). For these reasons, teachers pushed the district to consider developing more rigorous performance evaluations for all district non-instructional central office and classified staff (Shelby County Schools 2013). Research and development of a comprehensive, uniform non-instructional performance evaluation for the district began in June 2013. In 2013-2014, the district adopted a revised version of what the Memphis City Schools system had in place, but which had not been consistently applied to all employees, depending on the ability or willingness of certain district managers to comply with the evaluation program (Memphis City Schools 2012). In 2014, the district began a series of working groups of district employees and managers designed to provide feedback on a tentative non-instructional performance evaluation system and model rubric.<sup>3</sup> This is all worth noting only because it demonstrates how the entire effort of the district to create a non-instructional performance evaluation system, while distinct from the teacher assessment program adopted by the unified district, was essentially an effect of the district's focus on teacher evaluations. The initial interest in creating a non-instructional evaluation program as well as the steps taken along the way to research for, design and implement a comprehensive non-instructional performance evaluation system were wholly reliant on and guided by the district's ongoing work with teacher and principal evaluations (Shelby County Schools 2013).

---

<sup>3</sup> The author served as Shelby County Schools' research intern from June 2013 to April 2014, serving the Office of Performance Management and assisting with the design and implementation of a comprehensive non-instructional performance evaluation program for unified district staff, along with Jessica Lotz, Eric Linsy, Melissa McConnell, Kemmashela Smith and Valarie McCoy in the Office of Performance Management.



Longenecker and Ludwig, in their article on the implementation of performance appraisal systems discuss the importance of supervisors and administrators setting the example and being held to the same standards as subordinates (1990). They make their argument in the context of discussing the problem of inaccurate reporting of performance appraisal results, which will be discussed later in the study. They exhort firms considering designing, implementing or revising performance appraisals: “Organizations must provide leadership from above. Middle and lower level managers cannot be expected to provide accurate ratings if intentional inaccuracy is practiced higher in the organization” (1990, citing Arthur 1987). In much the same way, teachers and principals and lower-level staff cannot be expected to accept, support and devote themselves to the performance evaluation system used in Shelby County Schools if upper-level administrators and even school-based support personnel are not held accountable to high performance standards and to those they are meant to serve. Such a system creates friction, tension and jealousy between employees and is not conducive to the facilitation of a harmonious, team-like working environment in which each employee- central office administrator, school-based or other non-instructional personnel, principal and teacher alike- sees the value of his or her contributions to the overall achievement of the district’s most important goal, that of providing high quality teaching and increased academic achievement for all the public schoolchildren in the Shelby County Schools system. This transfer of policy goals from teacher evaluations to a new focus of the district, performance evaluations for the district’s non-instructional staff, is an example of how the teacher and principal evaluations designed by Shelby County Schools represent the most comprehensive framing of district goals, values and expectations. It is, if anything, a demonstration of how importance district officials see the work

of teacher evaluations and teacher effectiveness, that they have acknowledged the importance of aligning the instructional and non-instructional evaluation frameworks.

It emphasizes as well that, as a public school system, Shelby County Schools has sought to design and maintain systems which avail it of honest, accurate performance data which it can then share with the public and with invested stakeholders in order to seek and gain the community's support and involvement, secure valuable input for purposes of improvement, foster a sense of intimate involvement and inclusion in school decision making on the part of parents and community members and, ultimately, to continue to hold itself accountable to the public for the services it provides. In Shelby County, as in numerous districts around the country, a most obvious measure of the quality of the teaching services provided by public school systems has been seen to exist in evaluations of a district's teachers and, increasingly, of its principals. This, Shelby County Schools argues, is the most effective way to hold itself accountable for the quality of the necessary services they provide to the students residing in these districts (Klein 2011). Since the district's teacher evaluation system is designed to increase accountability, lessons from examining the merger context in which this system came about and has been significantly altered is valuable in considering the effectiveness and implications of teacher evaluation policy in Memphis and Shelby County.

Another element in the history of public education in the greater Memphis area must be considered before the story is complete, however. Over the course of the past summer, actions taken by residents of the incorporated municipalities of Shelby County have left the district poised to change its structure yet again beginning in the 2014-2015 school year. In the summer of 2013, referendums were held in each of Shelby County's six municipalities to determine

whether or not to establish their own independent municipal school districts. This decision was not instantaneous, however. It was the result of a long and tedious process pursued for almost two years by residents in the municipalities. It was not the first time that the schools and communities in the areas of Shelby County outside Memphis had tried to sever ties with the city and form a separate, special school district. Since 1993, the proposal to change the status of the Shelby County Schools district has been made in some form or another “by the Shelby County Schools board or its proponents” (Locker 2010). The proposals usually each attempted to effect the same result, namely that of transforming Shelby County Schools from a county school district, responsible for the education of all the students in a given county, into a special school district, bearing the responsibility to educate only those students within the confines of its borders and having authority to levy taxes to pay for school costs (Locker 2010). In the 2011 session, however, due to an influx of Republican lawmakers in the state legislature following state elections, proponents of giving Shelby County Schools special school district status hoped that they would have more success in repealing “an 18-year old state law banning” the establishment of new school districts (Locker 2010). Sensing the momentum, Representative Ron Lollar of Bartlett re-introduced his bill to carve out a special district for Shelby County Schools in the 2011 legislative session (Locker 2010).

*The Commercial Appeal* noted in November 2010, that chairman of the Shelby County Schools board David Pickler had “significantly improved...chances for success” as a result of the “legislature’s new Republican majority” (Silence 2010). Pickler was even quoted as saying that he was “willing to negotiate ‘virtually everything’” with city school system leaders “as long as the [county] school system” remained “safe from consolidation” (Silence 2010). By December

2010, following the vote of the Memphis City Schools board to surrender its charter and building on the growing anticipation of legislative success for law authorizing special district status for suburban schools, grassroots organizations in the suburbs surrounding Memphis began forming to promote and advocate for the maintaining of a separate county school system (Roberts 2010). Following the vote of the Memphis City Schools board to surrender its charter, the appeals to the state legislature for special school district status on the part of Shelby County Schools and community school advocates took on increased urgency (Silence and McMillin 2011). Other bills filed by supportive lawmakers at the beginning of the legislative session included those designed to delay or obstruct the merger referendum and implementation process, should the city-wide vote scheduled for March 8 be successful (Locker 2011). In February, the state legislature gave final approval to this legislation (Locker and Maki 2011). By September, Judge Mays had issued his consent decree, which he thought would bring “finality” to the issue of school consolidation (McMillin 2011). The decree outlined the settlement deal unanimously agreed to by both school boards “after nearly 20 hours of mediation sessions” (McMillin 2011).

On this basis, Mays argued that there should be “no dispute about the fairness of the agreement” and stated that the mediated settlement “prevents years of litigation and establishes the basis for cooperative solutions based on good public policy, rather than legal solutions imposed by the Court” (McMillin 2011). Nevertheless, the opponents of the merger were still not satisfied with the outcome. By November 2011, an organization based in the suburb of Bartlett, known as Better Bartlett Schools, began to host a series of meetings discussing the possibility of separating from the newly-created Shelby County Schools and creating a “future municipal school system” (Bailey 2011). Residents turned out to the town hall meetings hosted by the

group and argued in support for “a separate school system from the consolidated, countywide configuration on the horizon” (Bailey 2011). By January 2012, both Collierville and Germantown had begun the process toward the creation of municipal school districts in those suburbs, calling for public referenda to determine if the populations of the municipalities agreed with the creation of such a special district (*The Commercial Appeal* 2012). When asked why they were pursuing separate municipal school districts in the face of consolidation, residents of these suburbs cited their severe “lack of trust” of the newly-unified district and of how the ways in which Memphis City Schools’ running of its district would influence the administration and educational outcomes of the unified district (Bailey 2012). Residents noted that they were doubtful that they had a “viable voice” in calling attention to their concerns with a unified system and also felt that they had cause to “question whether the larger system will be responsive when things are needed” (Bailey 2012).

In May 2012, Tennessee Governor Bill Haslam signed into law a bill which “[paved] the way for Shelby County suburban cities to hold referendums this year on creating new municipal school districts” after the attempts to hold referenda by the municipalities were denied by the Shelby County Commission (Locker 2012). Shortly afterward, the six suburban municipalities of Memphis set a referendum date of August 2 on which to hold a vote on whether or not to move forward with the proposed creation of these suburban school districts (Bailey 2012). In response, the Shelby County Commission filed an injunction in federal court attempting to block the elections which could give way to municipal schools (Bailey, et al. 2012). The lawsuit argued that “race was a motivating factor behind the enactment of the Shelby County Municipal School Acts and that such Acts are purposefully discriminatory” (Bailey, et al. 2012). The county

commissioners filing the suit argued that, should the municipal school districts be created in Arlington, Bartlett, Collierville, Germantown, Lakeland and Millington, the districts would “be majority white” and would thus result in segregated education (Bailey, et al. 2012). The lawsuit also specifically “criticized” the law passed by the state legislature earlier in the spring allowing for the referendums to be held, saying that the legislation “can only be explained as an attempt to guarantee that Caucasian suburban children would never have to attend a school in the predominantly African-American Unified Shelby County School system- even for one year” (Bailey, et al. 2012).

On August 2, 2012, while the lawsuit was still pending in federal court, “voters in Shelby County’s six suburban municipalities gave landslide approval” to the creation of municipal school districts in an effort to avoid being a part of one consolidate Shelby County Schools system (Bailey, et al. 2012). In September, the lawsuit instigated by the Shelby County Commission alleging that the referendums and the municipal school districts which would exist as a result were discriminatory and unconstitutional, was taken to a trial (McMillin 2012). At issue were two primary questions involving “whether laws passed in 2011 and 2012 lifting the statewide ban on the creation of new municipal or special school districts violated the Tennessee Constitution by illegally applying new laws to Shelby County that do not affect other counties” and also whether or not the state requirement that, throughout the merger and municipal referendums, “all rights and benefits for any MCS or SCS teachers” would be guaranteed by the state and remain the same as they were before the merger and referendums took place (McMillin 2012). In November, U.S. District Court Judge Mays issued a ruling on the constitutionality of the referendums, ruling them unconstitutional, voiding their results and temporarily halting the

efforts of those in the suburbs attempting to secure education apart from the gradually developing unified school district (McMillin 2012). According to the ruling, the “legislative debate” over the law demonstrated that “Republican state legislators intended for it to apply only to Shelby County” (McMillin 2012). The problem with this arises in that this made “the law local in effect and unconstitutional since it did not include a provision gaining countywide approval,” which was necessary (McMillin 2012).

Former chairman of the board of Shelby County Schools David Pickler began to appeal to residents in the suburbs surrounding Memphis to work together with city residents to craft a school system that was beneficial for all students involved. He stated: “I think the door is wide open to build a structure that can serve all interests and be mindful that the judge has left open the creation of special municipal school districts, and we’re still dealing with a world in which there are many possibilities....If we continue to attempt to divide and conquer, we may find we have lost an opportunity to dictate our future and realize self determination” (McMillin 2012). Eventually, advocates of municipal school districts began to concede that the 2013-2014 school year would most probably feature a unified, countywide public school system (McMillin 2013). Collierville mayor Stan Joyner was quoted in *The Commercial Appeal* as acknowledging that, “Nobody has asked me directly, but if they did, I would tell them that it looks like we will be part of that unified school system for at least one year” (McMillin 2013). Proponents of independent, municipal school districts, however, were quick to insist that they were looking forward to the 2014-2015 school year during which they could establish and implement municipal school districts, expressing a desire to revise legislative proposals to ensure that the constitutionality of these laws would not be challenged (McMillin 2013). By February of 2013, representatives of

lobbying and advocacy groups dedicated to the establishment of municipal school districts in Shelby County, were meeting with lawmakers in Nashville to redesign their proposals and talk about how to craft laws which would allow the creation of special municipal school districts statewide so as to avoid the constitutionality issue of having the law only apply to Shelby County (Bailey and Locker 2013).

In April 2013, Governor Haslam signed into effect the laws that were the result of these talks with suburban officials, authored by State Representative Curry Todd and Senator Mark Norris, both of whom represent Collierville (Locker 2013). The law provided that, should voters once again approve measures to establish independent municipal school districts in Shelby County, the municipalities would still have “to follow the law on the process and requirements for new school systems and be approved by the state commissioner of education” (Locker 2013). The law also set the date for the start of operations for any new municipal school districts as August 1, 2014, guaranteeing that, at least for the 2013-2014 school year, all Shelby County public school students would be educated in one unified district (Locker 2013). Moreover the law took special note to protect the rights and privileges of teachers who might move from system to system and ensured that these would be guaranteed as constant by the state (Locker 2013). On July 16, 2013, voters in all six suburbs of Memphis approved measures aimed at creating special municipal school districts in each of the municipalities. As *The Commercial Appeal* reported, “approval numbers ranged from a high of 94 percent in Collierville to a low of almost 74 percent in Millington” (Garlington 2013). Community members and residents were ecstatic. One woman noted, “I’m extremely excited. As parents, we just want the opportunity to have a say in our children’s education system” (Garlington 2013). As of August 2013, school



board elections are being held for the second time, but this time with the understanding on the part of both suburban parents and activists and also unified school district administrators that the votes would actually be upheld (Garlington 2013).

Thus, over the course of only half a month, the schools in Shelby County experienced a remarkable amount of transformation. On July 1, 2013, the more than two year-long consolidation process was concluded, uniting for the first time the governance of all public education services in Shelby County. A little over two weeks later, the municipalities, which had constituted much of what was formerly part of the old Shelby County Schools system, voted to break away from the unified district. This move, which many conclude will be upheld as constitutional under the revised legislation, will most likely have the effect of largely returning many of the formerly SCS-run schools in the suburbs to the administration of a separate school district after this school year. Administrators at unified district's central office predict that the vote will effectively mean that school governance will look much like it did before 2010, prior to the Memphis board's dissolution of its charter and the long and difficult consolidation process. School board elections for each of the six individual boards are set for November 7, 2013, and each of the municipalities is currently in the process of studying how it will seek to create a separate school district by August 2014 (Bailey 2013). While the Shelby County Commission has "challenged the suburbs' pursuit of schools in federal court," many expect the referendum to be approved (Bailey 2013). At this point, most of the negotiations being undertaken by the unified district with each of the municipalities focus on how best to handle the breakaway, with representatives from each side discussing "attendance zones and whether the cities will be able to use the county school buildings with their respective boundaries" (Bailey 2013). The only

formerly SCS-run areas which will remain under the control of the unified district are those unincorporated areas of Shelby County which lie outside both the corporate limits of the City of Memphis and of the municipalities. In many ways, the process begun in December 2010, which saw the largest merger of any two school districts in American history, has been completely reversed. This context of transition and uncertainty on the part of community members, parents, students, teachers, administrators and policymakers is fundamental to an analysis of teacher evaluation policy in Memphis and in Shelby County.

As a result of the consolidation, old chains of command and administrative hierarchies have been disrupted and lines of accountability and supervisory authority have, in many cases had to be restructured and altered. In many ways, this manifested itself in the questioning of old norms and manners of doing things formerly held to by both districts. This resulted in a serious rethinking and reformulation of district goals and priorities and new ideas about how best to meet these priorities, all tailored to the specific and unique needs of the unified district. This has had direct implications on the ongoing implementation of the teacher effectiveness work. Of perhaps most immediate concern was deciding which evaluation system to utilize for the 2013-2014 school district. It was decided fairly early on by the Transition Planning Commission that Memphis' form of teacher evaluations would be adopted by the unified district. This has added an extra layer of complexity to the story of the development of teacher evaluation policy in Memphis to consider that teachers and administrators from the old Shelby County Schools system were first introduced to and subject to being evaluated by the framework adopted by Memphis City Schools with the help and assistance of the Gates Foundation in the 2013-2014 academic year. Their reactions, as teachers who are completely new to the system and who,

while working for Shelby County Schools, operated under a teacher evaluation system much like those adopted in the late 1990's and early 2000's which rated teachers on the basis of a more simplified checklist. This is not to say that policies and approaches used in the legacy Shelby County Schools system did not influence teacher evaluations as they now occur in the unified district. Much of the system of professional development and the actual rating and scoring process, conducted through a highly-detailed online system, have been influenced by old Shelby County Schools approaches and many of the county's district personnel now work in the Office of Talent Management administering these programs. Still, it must be noted that the bulk of the process and approach to teacher evaluations as they now occur in the unified district, along with the values and guiding policy inclinations behind the systems, are taken primarily from the work conducted in Memphis City Schools before the merger.

### **Teacher Evaluations in Memphis and the State of Tennessee: A Policy History**

In March 2010, Tennessee was awarded \$501 million in federal grants to carry out a program of school reform which it had outlined in its application for the Department of Education's Race to the Top competition and initiative (Tennessee Department of Education 2010). Tennessee and Delaware were the only two states selected as winners in the first round of Race to the Top and were selected from a pool of over 40 applications from states and from the District of Columbia and were declared winners after being chosen from among the competition's 16 finalists (Tennessee Department of Education 2010). In its application, Tennessee specifically explained its aggressive plans to improve teacher and principal

evaluation, use data to inform instructional decisions, and turn around [its] lowest-performing schools” (Tennessee Department of Education 2010). It also had already invested itself in the work of education reform, having “put in place strong laws and policies to support [its] reform efforts” (Tennessee Department of Education 2010). At the time of the award, Tennessee was determined by peer reviewers to contain one of the most impressive dedications to improving educational opportunities for its children. This was because of the evidence of the “commitment to reform from key stakeholders, including elected officials, teacher’s union leaders, and business leaders” (Tennessee Department of Education 2010). In addition, Tennessee and Delaware were some of the only states in which “all school districts committed to implementing Race to the Top reforms” (Tennessee Department of Education 2010). Perhaps the most notable part of its entire reform agenda, especially in the eyes of national evaluators, was Tennessee’s new system of teacher evaluations.

Passed in January 2010 and a crucial part of what the Department of Education considered to be Tennessee’s already-existing commitment to bold, unprecedented education reform, the new system allowed for “students’ academic performance” to be considered as “half of public school teachers’ job evaluations” for the first time in Tennessee history (Locker 2010). These moves were largely seen as heavily motivated by the state’s then-ongoing application to the Race to the Top program; they would, it was argued, “position the state to apply for...federal grants, in a competition with other states” (Locker 2010). The fifty percent of each teacher’s evaluation that would be based on student achievement data was comprised of two parts: Thirty-five percent would be “based on student gains on the testing done under the Tennessee Value-Added Assessment System” also known as TVAAS (Locker 2010). The other fifteen percent of

the student achievement portion of the evaluation would be based on “other empirical student performance...such as reading assessments for elementary teachers and college entrance tests...for high school teachers” (Locker 2010). This type of assessment, however, is only possible for those teachers whose students are required to undergo this type of testing. How the state evaluates teachers, such as special-education teachers, fine arts teachers and teachers whose grade levels are not given TVAAS tests was and is still a matter of contention for teachers and policymakers. Tennessee credited the policy, which was implemented in July 2011 under the title “First to the Top,” with the potential for “developing and improving great teachers and leaders in Tennessee classrooms” and with “placing a renewed focus on the classroom teacher and a more dedicated focus on encouraging student achievement” (Tennessee Department of Education 2011). The proposals were set to take effect during the 2011-2012 school year.

Tennessee began the policy by conducting a “field test” as it developed the “observation rubric” for the new teacher evaluation system (Tennessee Department of Education 2011). Through this system, Tennessee asked teachers “to tell policymakers what they need[ed] to promote and improve their classrooms” and what, in their opinions, would make an effective and accurate evaluation system (Tennessee Department of Education 2011). The state hired the National Institute for Excellence in Teaching or NIET to train “5,000 Tennessee teachers, principals and administrators in nearly 100 four-day training sessions across the state” with the express purpose of introducing the new evaluation system to the school districts and enabling them to gain an immediate understanding of how it worked (National Institute for Excellence in Teaching 2011). This was because the state decided to use the Institute’s “Teacher Advancement Program” or TAP standards as the basis of its new teacher evaluation model (National Institute

for Excellence in Teaching 2011). Also key to Tennessee's new teacher evaluation policy was its use of a new method of measuring teachers by their students' academic achievement, a method which now bears the state's name. The Tennessee Value Added Assessment System or TVAAS is a "statistical analysis of achievement data that reveals academic growth over time for students" (Tennessee Department of Education 2011). As opposed to assessing students against an absolute standard of where they ought to be performing, "TVAAS methodology follows the progress of individual students over time" (Tennessee Department of Education 2011). As a result of the use of this method, considered a significant positive development by many proponents of the teacher evaluation reform movement, "each student is compared to his/her own past performance" and "each student serves as his or her own 'control'" (Tennessee Department of Education 2011). Thus, student achievement, and, consequently, teacher effectiveness, in Tennessee was to be measured by how much a given student improved based on where he or she was positioned at the start of the year. Tennessee's use of this new method of measuring student data set its evaluation system apart from others being tried elsewhere in the nation and was a significant part of the state's later framing as a leader in educational reform, particularly in the area of teacher evaluation policy. The introduction of the TVAAS model for measuring student outcome data provided significant material for the continual national discourse on teacher evaluations and its merits are still being debated by scholars in the field. National discussion has weighed in on the method, mostly with uncertainty as to its effectiveness, however. As researcher Haggai Kuppermintz suggests: "Until a more complete case for TVAAS and the value added methodology has been developed, policy makers will be prudent to adhere to Ballou's (2002) admonition- '...those who look to value-added assessment as the solution to the problem

of educational accountability are likely to be disappointed. There are too many uncertainties and inequities to rely on such measures for high-stakes personnel decisions” (Kuppermintz 2003). He added that most of the support for TVAAS comes from “general statistical theories,” but that it is “only loosely aligned with relevant theories in education” (Kuppermintz 2003).

As the state implemented its new teacher evaluation system at the start of the 2011 fall semester, having gone through a summer of surveys, responding to feedback from teachers and administrators, and the training of the state’s vast network of employees serving in the public education system, it began to receive considerable pushback from teachers in the system who raised concerns as to whether or not its “methodical, time-consuming approach and 1-to-5 grading system” were actually effective measures of whether or not teachers were helping students learn (Nashville City Paper 2011). As the *Nashville City Paper* reported in the first few months of the program’s implementation, “As it turns out, stories of teachers critical of components of the system are plentiful” (Nashville City Paper 2011). Called TEAM (Teacher Education Acceleration Model), the final rubric was the work of the Tennessee Evaluation Advisory Committee “which included teachers, principals and other educators” (Nashville City Paper 2011). The group relied on surveys of teachers and administrators and “oversaw field tests that explored four models” (Nashville City Paper 2011). Eventually, the formula arrived at by the state was a system in which three components were considered- 50 percent was determined by “student achievement data” measured by the TVAAS model, 35 percent consisted of “in-class observations” and 15 percent was drawn from other tests and measures that calculated “student growth” (Nashville City Paper 2011). As principals began to fill classrooms to observe teachers in the classrooms for roughly ten minutes or so, teachers began petitioning the legislature to

reconsider the policy (Hardy 2011, Nashville City Paper 2011). While legislators promised to take teachers' concerns into consideration, it was understood that there would be no "scrapping or overhaul of the evaluation program" (Nashville City Paper 2011). It was argued that, "after all, it was this evaluation formula...that helped position Tennessee to score \$500 million in highly coveted federal Race to the Top funds" (Nashville City Paper 2011).

The new system was undoubtedly a radical departure from what had come before, whether for better or for worse. As Kevin Hardy of the *Chattanooga Times Free Press* reported in 2011, "Previously, a teacher was evaluated solely on an administrator's evaluation, which was given as little as once every five years. Changes in state law make this the first year that student achievement is used in teacher performance evaluations" (Hardy 2011). Even the way in which principals observed and reported on teacher performance and what they were attempting to find as a result of these observations had changed drastically with the law in 2010. For this, Tennessee was the recipient of much national attention in support of its policy proposals. In summer 2012, following the first year of Tennessee's implementation of its teacher evaluation program, U.S. Secretary of Education Arne Duncan authored an opinion editorial in *The Huffington Post* in support of Tennessee's efforts, titled "The Tennessee Story" (Duncan 2012). Duncan labeled the Tennessee experiment a vital success for the movement seeking to measure teacher performance by the performance of their students. Duncan analyzed Tennessee's policy, saying, "What are some of the takeaway lessons from Tennessee's experience? First, student growth can and should be one of a number of measures in evaluating the performance of teachers-- and it's important not to ignore a teacher's impact on student learning just because it is difficult to measure" (Duncan 2012). For Duncan, Tennessee was an important step forward in



the history of education reform: “As Tennessee has shown, our children, our teachers, and our country will be better off when school leaders and educators finally undertake the challenging task of creating a meaningful and useful system for supporting and evaluating our nation’s teachers” (Duncan 2012).

To confirm whether or not the program was actually a success, Tennessee hired a non-profit advocacy group, the State Collaborative on Reforming Education (SCORE) to spend time during spring semester 2012 to evaluate the strengths and weaknesses of the system and provide recommendations on where improvements might be made (Zelinski 2012). The SCORE report, released in June 2012, provided seven specific recommendations on how the state might improve its teacher evaluation program. Most notably, the report pinpointed the program’s failure to both sufficiently train or prepare teachers for evaluations and to give constructive feedback with the emphasis on allowing teachers to actually improve their work based on what they learn about their performance from evaluations (National Institute for Excellence in Teaching 2012). In addition, the report noted the importance of collaborating with teachers working in areas, subjects or grades where there were “gaps in the quantitative measure and some missing elements in the qualitative measure” (National Institute for Excellence in Teaching 2012). The report pointed out the unfairness in only ascertaining 50 percent of the information (the portion covered by the qualitative assessment by principals or administrators) needed for a teacher’s evaluation and leaving the rest of the evaluation unfulfilled if the teacher did not have “individual student growth, or value-added, data” available to them because of their field (National Institute for Excellence in Teaching 2012). The report suggested giving these teachers “the option of temporarily increasing the weighting of the qualitative portion of the

evaluation” (National Institute for Excellence in Teaching 2012). The report emphasized, above all else, the need for state administrators to embrace an attitude of continuous improvement, seeking out flaws in the system and openly and sincerely attempting to correct mistakes (National Institute for Excellence in Teaching 2012). In an interview following the release of the report, Tennessee Education Commissioner Kevin Huffman readily accepted this approach, saying that he was going to continue to seek the feedback of teachers and principals and that he recognized the system’s imperfections. Huffman stated, “One of the things we needed to communicate is that we would listen to teachers” especially concerning the ability of the program to be tweaked in successive years (Heitin 2012). Huffman remained adamant, however, on the point that the use of student achievement data would continue to be an important part of the evaluation’s rubric. He “emphasized that he will hold fast on incorporating school-wide performance goals into evaluations (an issue he’s received consistent pushback on) because it encourages all teachers to incorporate literacy and math into their instruction and because he believes it’s a better option than trying to create tests for every grade level and subject” (Heitin 2012). As one of the original focal points of Tennessee’s nationally-recognized education reforms, this emphasis on using student data to evaluate teacher effectiveness will almost certainly remain a key component of Tennessee’s teacher evaluation policy and will continue to spark statewide and nationwide discussion about its usefulness and its potential benefits and disadvantages.

As the state of Tennessee was pulling together its application for Race to the Top, Memphis City Schools, then a separate entity from the Shelby County school system, was gladly accepting the offer of the Bill and Melinda Gates Foundation to redesign its entire teaching

system, not only completely overhauling its previous teacher evaluation system, but also investing money in recruiting dynamic teachers and encouraging them to invest themselves in Memphis schools. The money was matched against funds contributed by other organizations, including the local teachers union, the Memphis Education Association (Roberts 2009). As contracts were signed between Memphis City Schools officials and leading officers of the Gates Foundation, Colleen Oliver, the Gates Foundation's senior program officer, said of Memphis' proposals: "We've watched the entire community come together around what we consider the most important challenge we face today...All the youth of every city, but particularly in the Memphis community, have a right to an effective teacher" (Roberts 2009). The funding represents the "largest gift to public schools in the city's history" and placed Memphis on the "front page of the nation" in terms of education reform (Roberts 2009).

According to the plan, \$1.9 million was to be used to video and evaluate the performance of the district's teachers in order to provide a starting point from which to determine workable measures for teacher effectiveness (Roberts 2009). With an added contribution of \$155 million from the Memphis Education Association, the city would completely redesign how it understands and evaluates teacher effectiveness through the use of elements including "performance, peer critique and subject knowledge" (Roberts 2009). Under the plan, current teachers would have the option of whether or not they wanted to participate, while "new hires" would be subject to these types of evaluations under their contracts (Roberts 2009). One of the criteria for acceptance into the Gates program was that the local teachers union had to be on board with the proposed changes and be willing to "negotiate changes in its contracts" that lined up with the broader goal of improving student achievement (Roberts 2009). On the state's

educational report card for 2009, “city schools received D’s and F’s” (Robert 2009). The plan to combat this, as outlined in the district’s application to the Bill and Melinda Gates Foundation included efforts to redefine teacher effectiveness and create a model for evaluations that were based upon this definition (Memphis City Schools 2010). In addition, the city planned to use the Gates Foundation money as well as grants from local organizations to raise teacher salaries and attract highly-qualified new teachers. As the *Commercial Appeal* reported in 2009, “The district will unilaterally raise salaries and pay the top tier of teachers salaries approaching six figures” (Roberts 2009).

In the report that outlined Memphis’ new Teacher Effectiveness Initiative, specifics are provided which clarify questions about how the district measures concepts such as teacher effectiveness. The report identifies the district’s definition of teacher effectiveness as the “Teacher Effectiveness Measure” or TEM and lists the “four components” of this measure (Memphis City Schools 2010). According to the Teacher Effectiveness Measure, teachers are evaluated as effective or ineffective on the basis of “growth in student learning,” “observation of practice,” “stakeholder perceptions,” and “teacher content knowledge and pedagogy” (Memphis City Schools 2010). The report also outlines the district’s steps to reform its teacher evaluation system to align itself with these standards. It states: “[Memphis City Schools] has improved the teacher evaluation process to provide a more fair and objective assessment of teacher performance. Our expectation is that an improved evaluation process will place us in a better position to individualize professional development experiences for teachers, create differentiated career roles based on performance, and establish a performance-based compensation schedule” (Memphis City Schools 2010). Memphis City Schools recognized its weaknesses and

the need to completely reform its teacher evaluation system, saying, “We currently operate within a framework that marginally identifies and responds to the strengths and weaknesses of our teaching corps” (Memphis City Schools 2010). The report recognized great room for improvement in terms of adjusting its teacher evaluation system to be more meaningful to teachers in terms of their understanding how it operates and more useful to administrators in terms of providing relevant information that actually helps them understand a teacher’s effectiveness in the classroom. The system of teacher evaluations in place at the time was recognized as “the source of anxiety and frustration for teachers and administrators” (Memphis City Schools 2010). Like the other districts in Tennessee at the time (prior to reforms made by the state on the basis of its application for Race to the Top funds), Memphis City Schools evaluated pre-tenure teachers annually and tenured teachers every five years (Memphis City Schools 2010). The actual evaluation process, described in the report as “an ambiguous series of events that are seemingly inconsequential,” consisted almost solely of a principals’ observation of a teacher in his or her classroom during class time and the principal’s gathering of “collateral,” consisting of “lesson plans, professional development plans, etc.” (Memphis City Schools 2010). Most importantly, the report concluded that “it is unclear how well the evaluation process lends itself to a valid and reliable view of teacher and student performance” (Memphis City Schools 2010). Such a system, the report argued, was in desperate need of reform since its basic success in achieving the purpose it was designed to accomplish was in question.

From the beginning of its reform efforts and its search for assistance from outside organizations like the Bill and Melinda Gates Foundation, Memphis City Schools attempted to include and listen to the voices of teachers as much as possible. In assessing the merits of both

the old evaluation system and the TEI, the city school system issued numerous surveys to teachers and received large levels of response (Memphis City Schools 2010). In addition, the city's acceptance into the Gates Foundation's program, which strongly emphasized a functioning, positive relationship of collaboration with the local teachers union on any reform proposals involving teachers, is strong evidence of its desire to work with teachers in its reform efforts (Memphis City Schools 2010). Regarding the level of cooperation in the redesign process on the part of teachers, the report argued that "teachers are interested in the TEI and are willing to give input into aspects of the reform that directly affect them" (Memphis City Schools 2010). It also added, "Teachers fundamentally agree that the current evaluation system needs improvement" (Memphis City Schools 2010). The report consistently acknowledged the reasonableness of teachers' problems with the "current evaluation tool" as a "vague and subjective indication of the teachers' performance that yields minimal understanding of what is needed to advance teachers to greater levels of proficiency" and attempts to ensure that the values that teachers seek in a reformed teacher evaluation system are included in the Teacher Effectiveness Initiative (Memphis City Schools 2010). This is perhaps why David Hill, one of the chief architects of the TEI called it "a much-deserved, though belated, compliment to the teaching profession" (Hill 2011).

In the report, Memphis City Schools outlined what it believed to be the justification for reform of the teacher evaluation system: "empirical evidence of the limitations of our current evaluation system" (Memphis City Schools 2010). The evidence came from the results of 1,400 pre-tenure teacher evaluations (Memphis City Schools 2010). The report noted that these results "reveal several interesting points about the quality and pattern of how evaluations are conducted

across the district” (Memphis City Schools 2010). It found that “of all ratings given to teachers, 90% fall into the ‘B’ or ‘C’ categories” and pointed out that “we shouldn’t have all of these low-performing schools if we have all of these high-performing teachers” (Memphis City Schools 2010). Instead, the report had great difficulty in even finding any sort of correlation between teacher performance evaluations and the grades that various schools received. The report stated that “there are no statistical differences between average performance on evaluations for high priority and good standing schools” (Memphis City Schools 2010). To conclude, the report resolved that “teacher evaluations should be a true representation of teacher performance and translate into quality professional development experiences for teachers across the district” (Memphis City Schools 2010).

Developments in recent years to the educational system in Memphis have affected the implementation of the Teacher Effectiveness Initiative as conceived of in its original report. Of these, perhaps the most significant is that of the school systems merger between the former Memphis City Schools and Shelby County Schools, as previously discussed. The answer to the question of how the merger has and will affect the Teacher Effectiveness Initiative and efforts to reform the teacher evaluation system in Memphis remains to be seen. Still, Memphis’ attempts to reform its teacher evaluation policy according to the standards of the Teacher Effectiveness Initiative are monumental in terms of their influence in both the state and the nation. In an article in the *Commercial Appeal*, David Hill notes that, “The TEI even served as a cornerstone in Tennessee’s successful \$500 million ‘Race to the Top’ application” (Hill 2011). He praises the potential that the TEI offers to Memphis students, saying, “Our community has before it a truly once-in-a-lifetime opportunity. When carefully implemented, the TEI will help ensure that all

students in our community are blessed with teachers who are as effective as today's best ones" (Hill 2011). Hill argues that, at a time "when approximately 4 percent of our city's public school students become 'college-ready' according to the ACT," the "TEI needs the strong and urgent support of everybody in this community" (Hill 2011). In Memphis, while the situation is still tentative because of the merger efforts, this backing of the entire community certainly exists, even if disparate portions of the community cannot agree on how to best implement reform. The recognition, on the part of the whole community, that the system is in desperate need of reform and the collective resolve to work together, even amidst controversy and strong disagreement, to find the ultimate solution, is perhaps what has drawn so much national attention to Memphis. As a center for reform-mindedness in education, with a willingness to engage in radical policymaking out of a desire to bring about a drastic turnaround in results, Memphis has tasked itself with transforming teacher evaluation policy and has framed this as one of the most important elements in its strategy for overall education reform.

### **Review of Relevant Literature**

The History of Performance Evaluations. Harvard's Robert Behn states, "everyone is measuring performance," noting the long-lived trend in the use of performance evaluation systems in a variety of contexts (2003). Much of the literature examined gives evidence of a comprehensive history of the use performance appraisal systems in private corporate settings and their eventual transfer to public sector organizations. Their use in urban public school systems has focused primarily on the evaluation of teachers, principals and upper-level instructional



personnel. This history of performance appraisal systems and their eventual significance for public education and for teachers, as it can be traced throughout the literature, not only provides an interesting backdrop to the research conducted in this study, but is worthwhile for the purpose of shedding light on some of the assumptions that surround both the use of performance appraisal systems in general and, in particular, their application in public education settings.

Even as early as the 1970's, such systems were described as being "frequently used in organizations as a basis for *administrative decisions* such as employee promotion, transfer, and allocation of financial rewards" (Brumback and Vincent 1970, DeCotiis and Petit 1978). During this same period, the United States federal government was already investing resources "to start developing policies and procedures on performance management," a process that gained significant traction with the U.S. Office of Personnel Management and which served as "the cornerstone of the Civil Service Reform Act enacted in 1978" (Brumback 2010). Even before this, the general outlines of ideas about performance management and its importance were present, albeit in very primitive stages, in Deming, Juran and Crosby's theories about leadership and business management, often collectively referred to as the "total quality management" theories. These were developed in Japan after World War II and increasingly applied in the United States in corporate settings in the 1950's as the nation's economy became less oriented toward manufacturing and increasingly more focused on providing services (Dwivedi 2001,

Besterfield, et al. 2003, Howard 2000, Rose 2005).<sup>4</sup> In the 1970's and 1980's significant contributions to the growing interest in using programming to assess and improve individual and organizational performance came in the form of the "excellence literature" which researchers describe as a "wave of best-selling non-fiction books that focus on organizational life and offer prescriptions for success" (Kee and Black 1985, see Hickman and Silva 1984, Grace 1984, Kanter 1983, Naisbett 1982, and Toffler 1980). The literature in this genre focused primarily on studying and understanding the key components of private corporate success by attempting to derive general, fundamental principles from the practical lessons to be learned from the structures and successful behavior of American companies and private enterprise. For example, significant texts cited by Kee and Black which promoted the study of corporations and how to achieve corporate excellence include such titles as *In Search of Excellence: Lessons from America's Best Run Companies* and *Corporate Cultures: The Rights and Rituals of Corporate Life* (1985, Peters and Waterman 1982, Deal and Kennedy 1982).

---

<sup>4</sup> In his article *Developing a Culture of High Performance: Some Research Findings and Experiences* published in July 2001 in the *Indian Journal of Industrial Relations*, R.S. Dwivedi expounds on the total quality management theories developed shortly after World War II. Edward Deming's theories of effective management, which valued highly effective leadership gained prominence in Japan shortly after World War II and was implemented widely. Joseph Juran, an American engineer and management consultant, also went to Japan as well after the war and popularized a management strategy there that emphasized corporate excellence through the use of effective management techniques. Philip Crosby's theories also relied heavily on the importance of management developing the values and goals of the organization and guiding it to its eventual success. Roughly two decades after the ideas of total quality management theory were first produced and became popular in Japan, a trend of American business leaders visiting Japan to witness the application of these theories gradually developed. Dwivedi describes the total quality management theory focuses on several components: "(a) involvement of engineering, production, quality assurance, marketing and human resources in preplanning during the product development stage; (b) *training people for high quality*; (c) *development of high performance strategies among employees*; (d) discussion and resolution of quality issues by employees; (e) ensuring defect-free parts by working with suppliers; and (f) adopting the concept of continuous improvement with the goal of zero defect" (2001, emphasis added). See Besterfield, et al. (2003), Howard (2000), and Rose (2005) for more information about the propagation of these prominent, early ideas about the importance of personnel quality which led to, at least in large part, the beginning of the trend in the private sector to emphasize employee performance and, ultimately, to attempt to measure it and make management decisions based on the data.

Ideas popularized by these authors included the notion that employees must understand a company's vision and how their individual work contributes to that vision as well as the importance of keeping company structures and operations as simple as possible (Kee and Black 1985). As Kee and Black identify, however, two particular concepts consistently stressed in these books were of particular importance and are of special significance to this report. They note that, of all the ideas stemming from excellence literature, "the belief that an organization's success is tied to its people...[who] need to succeed, to be winners, to feel that they contribute" has continued to have ramifications in both the public and private sectors to this day (1985). In addition, excellence literature authors tended to adhere to the conventional wisdom that "successful organizations measure their success" (Kee and Black 1985). In private corporations, especially those who ranked highest in output and productivity, there was an "insistence that performance be reported" and a "consistent use of such information" (Kee and Black 1985). The consensus was that America's private companies and firms were the standard-bearers and exemplars of efficiency and effectiveness in every capacity (Kee and Black 1985). These firms, according to excellence literature authors, operated on a set of key principles: That *human* resources were to be understood as mines of potential and capacity, that releasing this potential (i.e., enabling an employee to identify with the value of and feel the satisfaction in his or her work for an organization) was the key to any organization's success and that a system of measurement of human performance which could accurately identify weaknesses and propose improvements was essential to enabling employees to meet their full potential and thus bring about the achievement of an organization's objectives (Kee and Black 1985).

### The Introduction of Performance Evaluation in Government Institutions.

Soon afterward, some scholars began proposing the adoption of this ideology and approach by government agencies and argued that the public sector be more strongly orientated toward the ideas and practice of corporate excellence as described in the literature above (Hatry 1980, Sink, et al. 1984). In addition, during this time period, it was estimated that some roughly “92% of all organizations in the U.S. use some form of this process”; the use of performance evaluations was considered “a hot topic...for both the academician as well as the practitioner as organizations attempt to use the formal appraisal process as a critical human resource management tool” (Longenecker and Ludwig 1990, citing Locher and Teel 1988, Landy and Farr 1983). While some authors attempted to call attention to the importance of not “glossing over fundamental differences between tax-supported entities and market-supported organizations,” suggesting that such an error may “do both sectors a disservice,” the movement gained increasing popularity (Wildavsky 1964, Leone 1984, Kee and Black 1985). Studies such as those conducted by Schiflett and Zey attempted to point out the fundamental values and structural differences between what they termed “product producing organizations (PPOs)” and “public human service organizations (PSOs)” (1990). Nevertheless, a significant number of authors continued to assert that “there was one best way to organize” regardless of whether organizations were oriented toward public service, on the one hand, or product creation, on the other (Schiflett and Zey, citing Zedlewski 1979, Musolf and Seidman 1980, Cutt 1982, Gold 1982). Research by other authors argued that, due to the boundaries between the two types of organizations becoming blurred, the two types of organizations should not be seen as separate types, thus

implying that organizational strategies which were heavily utilized in the private sector (i.e., formal performance appraisal systems) might be applied with equal success to public sector agencies and organizations (Schiflett and Zey, citing Murray 1975, Shostack 1977, Bozeman 1987). As Schiflett and Zey note: “Accordingly, [these authors] support applying the business administrative practices of PPOs [private sector entities] to PSOs [public sector and non-profit entities]” (1990). Despite these concerns the trend continued to grow. By the 1990’s, the phenomenon of government agencies increasingly relying on extensive data-gathering efforts to measure organizational performance to assess and improve productive output became a popular subject for research and was increasingly utilized by government agencies, most notably those of state and local governments (Behn 2003). Not surprisingly, one of the most widely-implemented of these types of measures were formal employee performance evaluations (Behn 2003). This makes sense because, as some of the authors note, “A major part of government budgets pay for the people who do the public’s business. They are the teachers in public schools, professors at institutions of higher education, highway patrol troopers, soldiers, social workers, prison guards, engineers, tax collectors, and wildlife officers” (Kee and Black 1985).

As a result, performance evaluations began to gain prominence and became widely used as measures of the effectiveness of governmental organizations. Blodgett cites formal performance evaluation systems as “the hottest topic in government today” (Behn 2003, citing Blodgett 1996). Other authors refer to the “explosion in the use of performance indicators,” referring to such tools as having become “ubiquitous in the public sector” (Smith 1990). The link between the augmentation of public sector use of performance measures and the ideas and principles of private corporations and the corporate “excellence literature” seems to be firmly

established. The emphasis, scholars note, was relatively new to the public sector in the 1970's and 1980's and "draws on private sector-derived accounting and management technologies for the pursuit of public sector efficiency" (Padovani et al. 2010, Lapsley 2009). Now, it seems that governments increasingly took to publishing reports of performance evaluations as high-level indicators of their effectiveness and, it was noted, most industrialized Western nations were moving toward "the development of measurement systems which enable comparison of similar activities across a number of areas" and which "help to establish a performance-based culture in the public sector" (Behn 2003, citing Murphey 1999 and Kouzmin 1999). By 1993, the movement had gained enough national prominence such that Congress passed the Government Performance and Results Act (Behn 2003). In a report conducted under the newly-passed law, the National Academy of Public Administration (NAPA) declared: "Performance measurement of program outputs and outcomes provides important, if not vital, information on current program status and how much progress is being made toward important program goals" (NAPA 1994). In this context, performance evaluations in the public sector came to embody larger civic values and, it was argued, served an essential democratic purpose which their more traditional corporate counterparts did not: more readily informing citizens and thus enabling them to hold their governments accountable (de Lancer Hulnes and Holzer 2001). Wholey and Newcomer claim that "the current focus on performance measurement at all levels of government and in nonprofit organizations reflects citizen demands for evidence of program effectiveness that have been made around the world" (1997). Osborne and Plastrik asserted that the growing tendency on the part of governments to use performance evaluations "enables officials to hold organizations accountable and to introduce consequences for performance. It helps citizens and customers

judge the value that government creates for them” (2000). More practically, authors like Wholey and Hatry noted that “performance monitoring systems are beginning to be used in budget formulation and resource allocation, employee motivation, performance contracting, improving government services and improving communications between citizens and government” (1992).

In 1999, a survey of municipalities in the United States found that “some 38 percent of the [695] respondents indicate that their cities use performance measures” (Behn 2003, citing Poister and Streib 1999). The authors of the study concluded, however, that this was “a significantly lower percentage than reported by some of the earlier surveys” and other studies reported that municipal governments nationwide had a “meager record” of using performance measurement programs as a component of their regular organizational practice (Behn 2003, citing Poister and Streib 1999, Ammons 1992). On top of this, Behn qualifies his own assessment of the trend of increasing usage of performance evaluation measures on the part of state and local governments, saying, “And, of course, people who *report* they are measuring performance may not really be *using* these measures for any real purpose. Joyce notes there is ‘little evidence that performance information is actually used in the process of making budget decisions’” (2003, Joyce 1997). A number of scholars have identified that, while widely discussed and circulated, policies that actually implement and are based on performance measurement tactics are not as popular or as frequently used as has been purported. These scholars call for a distinction to be made “between rhetoric and reality” in terms of the pervasive use of performance evaluations (Padovani et al. 2010, Bouckaert and Peters 2002, Chan and Gao 2009, Midwinter 2008, Streib and Poister 1999, Taylor 2009). Why such an exaggeration of the use of performance evaluations

seems to have been assumed in some of the literature remains an interesting question in and of itself.

While these considerations must be taken into account, it is important to note that scholars still feel comfortable identifying the use and acceptance of performance management techniques by public sector entities “with ever-increasing regularity” as a strong trend with identifiable consequences for policymaking (Bertelli and John 2010, Padovani et al. 2010). Often referred to as “the concept of New Public Management (NPM),” the movement has gained significant traction in Western countries and has come to be commonplace in its impact on approaches to policymaking as well as numerous already-implemented policies (Padovani et al. 2010). Even once the limitations in the use of performance measurement by the public sector are acknowledged, it is still worthwhile to note the significant effects this movement has had on governments, how they operate and how they are perceived (Bertelli and John 2010, Padovani et al. 2010, Behn 2003). In addition, some authors have noted “a resurgence” in the use of “performance management systems,” observing that they are often “an essential component of professional public management” (Padovani et al. 2010, Halachmi 2005a, 2005b, Moynihan 2009). For these reasons, the increased use of performance evaluations in public sector organizations is arguably a relevant phenomenon and worthy of study. The increased interest in improving government performance through applying private sector principles of organizational management and structuring has continued to develop into the present day, influencing other prominent public sector, government agency or civil service reform movements and inspiring interest in widely debated and popular practices such as decentralization and privatization (Boyne, et al. 2007, Batley and Larbi 2004, Pollitt and Bouckaert 2004).



Continuing in the tradition of excellence literature in adhering to private sector principles and applying these to public sector agency reforms, “many of these policies have focused attention on the internal characteristics of public organizations” in much the same way as concern with the organization, its structures, policies, leadership and human potential, animates private sector literature (Boyne, et al. 2007). Included in this trend is the continued emphasis placed on employee performance appraisals (Bertelli and John 2010). Interestingly, such increased scrutiny on the performance of government agencies and the resultant increased effort expended in measuring and improving this performance has coincided with an increase in the “criticism of public-sector service delivery” over roughly the last thirty years (Behn 2003). The cause of this is uncertain. Some suggest it is due to the fact that public sector agencies, having accepted and incorporated private sector mentalities regarding performance and approaches to improving it, are more easily compared to their private sector counterparts who are able to devote significant amounts of time and energy to researching and bettering “the process measures that customers value” (Behn 2003). Perhaps the increased study of public sector performance has merely identified shortcomings in governmental performance which have long existed, but which were largely undetected due to the lack of investigation into the subject previously. An alternative cause might be the constant emphasis currently placed on achievement and performance in the culture surrounding private and, now, public sector institutions and their efforts which could have perpetuated misdirected opinions and expectations on the part of the public about how government ought to perform and what purposes it ought to serve.

Regardless of why public dissatisfaction with the performance of government agencies began to grow, one thing was certain- during this period in which increasing emphasis was

placed on performance measurements for governments, the *way* in which governments were evaluated was transformed. Its ultimate effect was to “realign the focus of government accountability and performance analysis away from activities and process measures and toward results or outcomes” (Heinrich 2002). Heinrich states that, for example, “the number of trainees placed in jobs” in a career development and assistance program, was considered a much more rigorous measure of program effectiveness than “the number of persons trained” in its own right (Heinrich 2002). Such a transformation was the product of the reorientation of performance appraisal systems toward “program outcomes” as opposed to “inputs or outputs” (Heinrich 2002). As a result, some scholars have noted, “public administration finds itself in an era of government by performance management, which is reflected in the widespread assumption that management is a key determinant of performance, and that it is reasonable to expect managers to measurably improve organizational effectiveness” (Moynihan and Pandey 2005). These assumptions, the intellectual descendants of both long-lasting private sector corporate management principles and their application in the sphere of public sector agencies, loom large in virtually every major national conversation over significant areas of public policy (Moynihan and Pandey 2005).

#### The Introduction of the Concept of Performance Measurement in a Public Education Context.

One such area of policy in which personnel appraisal systems have attracted significant attention is that of education reform. The publication of *A Nation at Risk* in 1983 called attention to what was referred to as “a rising tide of mediocrity that threatens our very future as a Nation

and as a people” (The National Commission on Excellence in Education 1983). The “groundbreaking report,” as some authors refer to it, immediately caught the attention of policymakers and activists with its emphasis on student performance on standardized tests as indicators of the organizational performance or output of the American educational system (Klein 2011). Since 1983, education policy and its many waves of reforms have consistently attracted national attention and it has become one of the most widely-discussed subjects in terms of government efficiency and performance. It is not surprising then, that current education policy reflects the larger tendency on the part of governments to introduce private sector performance-oriented or results-based measurement programs into the gamut of policymaking. With the passage of the No Child Left Behind Act in 2001, federal, state and local education policymakers attempted to better identify and target those schools which were underperforming according to state measurements of students’ Adequate Yearly Progress (AYP) and to attempt to solve the problems in these schools (Balfanz, et al. 2007, Bertelli and John 2010). NCLB represented a significant shift in the structures around which education policy had been centered in the past and a reorganization of bureaucracy and hierarchy, but not necessarily a change in values or in fundamental ideas about whether or not assessment-based decision-making was a worthwhile endeavor (Fuller, et al. 2007). Rather, policymaking in education began to heavily utilize student performance data from state-administered standardized tests in the early 1980’s after the community was mobilized after the publication of *A Nation At Risk* (Fuller, et al. 2007). Such “systemic” reform, initially led by states, focused on tracking and improving institutional performance, using assessments of student progress through the AYP as measurements of this progress (Fuller, et al. 2007). Before such testing systems were instituted, the only sets of

comprehensive national data to which researchers had access in order to track student learning for the purposes of informing federal policymaking were the scores of students who had taken the Iowa Test for Basic Skills (ITBS) as well as national SAT scores, as researchers at the Congressional Budget Office learned when Congressional leaders asked them to track postwar student performance (Congressional Budget Office 1986). Even once this was determined, researchers were clear to qualify what they had found. They “emphasized that students taking the ITBS and the SAT were not representative of the nation’s children” (Fuller, et al. 2007). In response, scholars began to propose “a state-led model of organizational change that called on governors and school districts to sharpen what children should learn and then to align state tests to these transparent standards” (Fuller, et al. 2007, Cohen 1990, Smith and O’Day 1991).

The new system was refocused on measuring student achievement outcomes, a process conducted by state departments of education (Fuller, et al. 2007). Fuller, et al. specifically note the origin of these new approaches to education reform, stating: “These fresh elements of systemic reform were borrowed, in part, from the writings of popular management gurus who reported how centralized managers in leading leading companies were focusing their efforts on tracking the performance of local units while decentralizing the means by which site managers pursued greater efficiency and innovation” (2007). Peters and Waterman’s (1982) manuscript on private corporations as exemplars of excellent institutional performance, *In Search of Excellence: Lessons from America’s Best Run Companies*, is again cited as a reference (Fuller, et al. 2007). There are other noticeable similarities between the initial transfer to and use of performance appraisal by public sector systems in general and their gradual acceptance by the nation’s educational establishment as one of its most consuming policymaking objectives.

Almost as early as the use of performance measurement systems in public education settings were first proposed, such application of these principles was justified by ascribing to them both practical, organizational and also civic, political value (Fuller, et al. 2007).

Due to the dramatically increased attention paid to the performance of the nation's schools, government spending on education had significantly increased during the 1980's (Fuller, et al. 2007). Researchers note that "state spending on education had increased by 26% in real, inflation-adjusted dollars between 1980 and 1987" (Fuller, et al. 2007). Emphasizing measurements of student outcomes as the primary directors of policymaking priorities, it was argued, "would advance public accountability and greater political legitimacy" as well as meeting the practical organizational and decision making needs of school districts and state education departments (Fuller, et al. 2007). In his 1990 report proposing this new model of assessment, Michael Cohen of the National Governors Association argued that such reforms would be useful to the public and to those at work seeking to better the nation's education system. He stated that, as a result of the increased attention on and spending directed toward public schools, "policymakers are expressing increased concern over accountability, asking if investments made in previous years are paying off in terms of performance" (Cohen 1990).

The arguments were compelling and, by 1999, 48 states had "put in place statewide testing programs" (Fuller, et al. 2007, Goertz and Duffy 2001). Often, these tests were designed similar to exams administered by the National Assessment of Educational Progress (NAEP), which had been established in 1964 (Fuller, et al. 2007). The NAEP had used the public's newfound concern with the quality of educational offerings nationwide, particularly for minorities and economically disadvantaged students, to solidify its role as the recognized

assessor of national educational progress and the quality of student learning. Soon, the states began to follow suit, measuring through their newly-created exams “student levels of mastery (criterion referenced) rather than percentile rankings that compared one state’s scores to other states (norm referenced)” (Fuller et al. 2007). In this way, measurements of educational progress were transformed. States and the NAEP were attempting to judge educational performance of states and school districts by assessing their absolute progress on standardized tests against pre-determined “cut-points for proficient levels of performance” which represented where students *ought* to be performing on these newly-created tests, whereas previously they had been assessed through comparisons of their data with other performance results from around the country (Fuller et al. 2007). In this way, the already relatively new emphasis placed on test results as a measurement of the success of, not only state and district educational programs, but also, more profoundly, the academic success of students gained added importance as it began to be used as a “criterion-referenced” evaluation of learning (Fuller et al. 2007). Even before the No Child Left Behind Act (NCLB) was passed, the vast majority of states had already set up their own testing systems which were overseen and administered by their education department and had already begun using these tests to measure and report their data in these ways (Elmore, et al. 1996). The law primarily mandated what has already been recognized as a well-established national trend toward judging student progress by performance on standardized tests against legislated or prescribed benchmarks for what was to be considered “proficient.” What the law does represent is a clear distillation of the fundamental assumption on which these types of policies rest: That student performance at the levels prescribed does in fact represent true and authentic student learning at the basic, proficient and advanced levels (Fuller et al. 2007). In the laws of the NAEP

itself, its tests and the use of their scores for data-based funding and policymaking choices operates under the following assumption: “Students reaching this level [of proficiency] have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter” (Brown 2000). Identification of this underlying assumption of these types of policies is important for the background and context of this report, however, an assessment of its accuracy and its implications is beyond the scope of this study.

#### The Introduction of Teacher Performance Evaluations in the United States.

Immediately after the use of student performance data became widespread and was subsequently mandated by NCLB the question arose regarding what types of decisions ought to be based on such data. States quickly began assessing organizational or agency effectiveness, determining where best to deploy federal, state and local resources and, of foundational significance for this report, assessing the behavior, skills and performance of the sector’s personnel. The most obvious and, consequently, the most widely targeted group for data-based performance assessments were teachers (Cornett 1995). A large number of researchers, activists and policymakers have argued that the use of student performance data as a significant component of the performance evaluations of teachers and other school personnel is an essential component of successful education policy reform (Klein 2011). As some scholars have noted: “Recent educational reform efforts seek to employ standardized test score gains as a key policy instrument for holding educators and school systems accountable” (Kupermintz 2003). Here, as

with the use of other public sector performance appraisal systems, the value of public accountability of agencies and personnel factors as a crucial policy consideration factors heavily into discourse on this subject. Such policy ideals not only represent the “cornerstone of the No Child Left Behind legislation,” but also see their continued propagation and reformulation in the Obama administration’s \$4.35 billion Race to the Top grant program (Linn 2003, Weiss 2009). Of the four key policy components which the program expects applying states to emphasize in their proposals, two are of particular interest. States winning Race to the Top grants to improve their education systems are expected to rely on proposals aimed at better identifying, recruiting and retaining “effective teachers and principals” and to make such human resources decisions using “data to inform decisions” (Weiss 2009). It is in large part due to these considerations and emphases that Tennessee’s application to the Race to the Top Fund included, as one of, if not the, most significant policy proposals in its First to the Top program the student performance data-based evaluation of teachers and other instructional personnel through the use of its newly-devised Tennessee Value-Added Assessment System (TVAAS) (Finch 2012, Kupermintz 2003).

Nationwide the movement attracted both significant national support and also fierce criticism. Data-based teacher evaluation program advocates such as Michelle Rhee, former Chancellor of the District of Columbia’s public school system and Joel Klein, former Chancellor of New York City’s Department of Public Education have argued that, of all aspects of the American education system in need of repair, the growing movement to “analyze year-to-year student progress and tie it to individual teacher performance” is the most crucial element in effecting true education reform (Klein 2011). As a result, since roughly 1980, an increasingly interconnected network of those invested in the national conversation on education reform has



formed and has consistently pushed for the development and implementation of data-based teacher evaluations and other policies aimed at improving teacher effectiveness through professional development, attracting and retaining teachers deemed to be highly effective to some of the lowest performing schools in the nation, creating high-stakes for teachers by directly linking systems of bonuses and compensation increases to that teacher's yearly student performance data, dramatically reforming teacher tenure laws, and attempting to edge those who are determined to be "low-performing teachers" off of school district payrolls (Cornett 1995). State legislatures, governors and education departments, Congress, the president, the Department of Education, local boards of education and a host of non-profit and for-profit research and advocacy organizations centered around education reform have all played a vital role in contributing to the copious amount of attention paid to education reform in general and teacher evaluations in particular (Cornett 1995). In Tennessee and in other states which use the value-added data-based model of teacher assessment, the use of value-added data has come to hold great significance and is based off of, once again, key assumptions. Haggai Kupermintz identifies these assumptions in his analysis of the TVAAS-based system and trends, saying, "By modeling student progress over time, the argument goes, value-added analyses provide accurate and trustworthy quantitative measures of student learning. These measures, in turn, can be directly attributed to the professional efforts of individual educators and schools, thereby mitigating the 'many problems in assessment and measurement'" (2003, citing Sanders 2000). Tennessee's teacher evaluation system was, arguably, one of the most significant reasons for why it was one of the first two states along with Delaware to win Race to the Top grants in 2010 (Finch 2012, Tennessee Department of Education 2010). Due to the national attention garnered

by the award, Tennessee's use of the value-added model for teacher evaluation programs has sparked "impetus for the development of new teacher evaluation systems, using longitudinal analyses of test data" nationwide (Kupermintz 2003). It is, researchers have noted, currently the most influential value-added model" (Kupermintz 2003).

In national conversations, the subject of teacher evaluations has been given increased importance and has sometimes been framed as singularly the most important aspect of effective education policy by some authors (Koppich 2004, Klein 2011). Of all the individual initiatives that are tied to education reform, the vast majority of studying, thinking, writing and policymaking that has been done with respect to education has often primarily or solely focused on improving or overhauling teacher evaluation systems (Koppich 2004, Winters 2012). What has resulted has been a "flurry of laws" and an "unprecedented wave of state-teacher evaluation reform across the country" (Cornett 1995, McGuinn 2012). The arguments made in favor of such rigorous teacher performance assessment models are often based on assessments of the performance of what is seen as the nation's poorly performing educational system and assertions that these seemingly "simple" and "common sense" solutions are essential to reversing this trend (Cornett 1995, Koppich 2004, Klein 2011). As researcher Julia Koppich argues, "Research has shown what common sense would suggest: Children with high quality teachers are able to make continuous progress. Those with less effective teachers often struggle, often not successfully to keep up" (2004). Marcus Winters, a senior fellow at the Manhattan Policy Institute, also makes this argument in his report, saying, "Several studies show that to which teacher a student is assigned makes a huge difference in determining how much that child will learn in a given school year. And many of the benefits of good teachers are more long-lasting" (Winters 2012). In

the popular media the movement has been touted as the end-all education reform effort as well. One scholar notes: “In *Waiting for Superman*, the 2010 documentary that describes the failure of American public education, several children and their families, along with educators like Geoffrey Canada and philanthropists like Bill Gates, drive home the argument that the key to school reform lies in improving the competence and skills of individual teachers” (Leana 2011). The early conversation about teacher evaluations, once academic and research-oriented has now exploded into a national movement involving even national foundations, “with the largest and most powerful, the Bill and Melinda Gates Foundation, providing hundreds of millions of dollars in funding to initiatives for improving teacher competence and accountability” (Leana 2011). Significantly, Memphis was one of the recipients of such a grant from the Gates Foundation specifically aimed at creating its new and revamped Teacher Effectiveness Measure (TEM), now in its third year of development (Roberts 2009). As it was argued, “All the youth of every city, but particularly in the Memphis community, have a right to an effective teacher,” and the sure sign of whether or not this right was being protected was a teacher’s performance as measured by a rigorous, value-added data-based teacher evaluation program (Roberts 2009).

Acceptance of the values, reasoning and policymaking preferences of this movement and the almost sole importance it places on teacher evaluations, however, is not at all universal. As scholars have pointed out, the collective effort to transform systems of teacher evaluations “has turned out to be one of the most controversial questions facing reformers” in the entire realm of education policy (Winters 2012). Other have argued that “there is no topic on which opinion varies so markedly as that of the validity of basing teacher effectiveness on student learning” and also that the proposals have resulted in mixed reactions on the part of any evaluation system’s

most crucial stakeholders- teachers (Shrinkfield and Stufflebeam 1995, Kupermintz 2003). They state that “persistent substantive and methodological shortcomings have contributed to ‘teacher skepticism and growing criticism of attempts to link learning gains to teacher work’” (Kupermintz 2003, citing Millman and Schalock 1997). Opponents of the burgeoning movement raise objections to, most notably, underlying assumptions and ideas about *how* teacher effectiveness ought to be defined and judged and the particular metrics that have been created to assess individual teachers as they measure up against these preconceptions of effectiveness (Rothstein 2010). They also express concern that the emphasis currently placed on teacher evaluations in national discourse on education reform has minimized attention being placed on factors which, they argue, are just as crucial to understanding the nature of the public education system and thus must necessarily be incorporated in any meaningful effort to shape the success or failure of America’s public schools (Rothstein 2010). For these authors, the constant emphasis on and implications of combating so-called “incompetent teachers,” which are seen by teacher evaluation proponents as the “single most important factor” in improving students’ educational success is an oversimplification of a complex issue resting on problematic or unproven assumptions, does a disservice to those most invested in the ongoing work of delivering quality education to the nation’s most vulnerable and often results in policies which are either ineffective or which cause more problems than they solve (Rothstein 2010). For opponents of the movement like education historian Diane Ravitch and economist Bruce Baker of Rutgers University, there are too many problems with the links being made between the research on value-added data and its use in teacher performance appraisals and the policies being argued for on the basis of this research (Strauss 2012 and Baker 2012, citing Ravitch 2012). As Baker notes: “Just because

teacher [value-added] scores in a massive data set show variance does not mean that we can identify with any level of precision or accuracy which individual teachers...are 'good' and which are 'bad.' Therein exists one of the major fallacies of moving from large scale econometric analysis to micro-level human resource management” (Baker 2012). Despite the continuing debate over the effectiveness or fairness of value-added data-based teacher evaluation programs, there is little doubt that these policies, the research which preceded them and the discourse which has ensued has significantly effected how both scholars and the public think about and react to public education policy and constitutes one of the most widely-discussed issues of education reform.

#### The Purposes for Implementing Performance Evaluations.

The literature has produced varied opinions with respect to the purpose for which institutions ought to design performance evaluations. Robert Behn, in his article titled “Why Measure Performance? Different Purposes Require Different Measures,” notes how such a seemingly simple subject can actually become quite complex (2003). He asks: “What purpose-exactly- is a public manager attempting to achieve by measuring performance? Even for this narrower question, the answer isn't obvious” (Behn 2003).<sup>5</sup> He cites Theurer, who argues that “the intent of performance measures is to provide reliable and valid information on performance” (1998). For Behn, however, this explanation is unsatisfying and incomplete, saying, “but that hardly answers the question” (2003). The question remains as to how managers of agencies with

---

<sup>5</sup> Behn's article, published in the *Public Administration Review*, specifically speaks to performance evaluation systems designed for public or governmental organizations.

performance measurement programs are expected to use the data they produce. Thus, Behn argues: “All of this reliable and valid data about performance is of little use to public managers they lack a clear idea about how to use them or if the data are not appropriate for this particular use” (2003). He cites a number of purposes posited by different authors as to why performance evaluations are instituted and utilized in the fashion they are. Some authors cite the importance of using such programs (in a public sector context) to enhance government accountability either to its citizens at large or to the specific clients a particular program or agency is designed to serve as seen earlier in this discussion (Wholey and Newcomer 1997, Wholey and Hatry 1992, Ammons 1995, Osborne and Plastrik 2000, Kravchuk and Schack 1996). With such accountability data, write Plastrik and Osborne, citizens will be able to ultimately “judge the value that government creates for them” (2000). The implication is that such citizens will eventually utilize such data to decide whether or not their interests are being served by various policies and agencies and to react to this data using their influence and power, through voting, activism, advocacy or other avenues (Osborne and Plastrik 2000). Others see it as an important tool for organizations to share information or “evidence” (Wholey and Newcomer 1997) about strengths, weaknesses and organizational effectiveness with citizens or stakeholders (Kopczynski and Lombardo 1999, Wholey and Hatry 1992). Occasionally, the motivation for introducing such an evaluation system, while seen as for the purpose of holding the agency or program accountable, but from the perspective of legislatures, administrators or oversight agencies instead of citizens or consumers of the agency or program’s services (Wholey and Hatry 1992, Ammons 1995, Osborne and Plastrik 2000, Kravchuk and Schack 1996).

While measuring performance to hold organizations, usually contextualized as public sector organizations or government agencies, accountable to the public in order to encourage the organization to meet high standards for service delivery, there are other purposes for introducing performance evaluations that are described in the literature. Some are derived from the practices of private sector for-profit organizations who attempt to assess the firm's human capital for purposes of maximizing productivity, while others seem to be designed to meet concerns that are uniquely faced by public sector organizations with indirect relationships with diverse consumer populations who may not have voluntarily chosen to receive its services or seek its assistance (Kee and Black 1985). Such theorized purposes include aiding the organization in "budget formulation and resource allocation" (Wholey and Hatry 1992, Hatry, et al. 1990, Ammons 1995, Kravchuk and Schack 1996), in "employee motivation" (Wholey and Hatry 1992, Kopczynski and Lombardo 1999), and providing important information to facilitate efficient and flexible management decision making (Wholey and Newcomer 1997, Hatry, et al. 1990, Osborne and Plastrik 2000, Kravchuk and Shack 1996). Others purposes for such systems posited by some authors include "improving government services" and "program effectiveness" (Wholey and Newcomer 1997, Wholey and Hatry 1992), "organizational learning" (Kravchuk and Schack 1996), "compar[ing] performance among a subset of jurisdictions" or similar entities (Kopczynski and Lombardo 1999) and "solicit[ing] joint cooperation in improving future outcomes in respective communities" and building community buy-in or trust with regard to organizational programming (Kopczynski and Lombardo 1999).

Another discourse has ensued between authors over whether or not performance evaluation systems exist to inherently reaffirm managerial authority, enforce control over

subordinates within an organization's structure and re-establish the hierarchical system of authority and governance. According to a theory advanced by some scholars, "employee appraisal is an objective, rational, and systematic attempt on the part of the manager to accurately describe subordinate performance" in order to exert and maintain control over the organization and its employees (Longenecker and Ludwig 1990, Sherman, et al. 1988, Kravchuk and Schack 1996, Behn 2003). Other authors disagree, however, referencing "excellence literature" and instead argue that "managers have a variety of concerns that are clearly more pressing than simply generating brutally accurate and honest ratings" and emphasize the context of "organizational environments that place a high priority on getting results, on minimizing conflict, and ultimately, on survival" (Longenecker and Ludwig 1990, Longenecker 1989, Peters and Austin 1983). In this way motives of managers self-interest enter into the performance appraisal process, either to the employee's benefit or disadvantage, depending on whether it is in the manager's interest to inflate or deflate ratings of employee performance (Longenecker and Ludwig 1990, Buchholz 1989, Longenecker, et al. 1987). This begins to enter into the question of not *why* performance evaluations are designed, but rather *how* they are implemented and enacted from manager to manager. This is a different question altogether.

The issue of defining, setting, adjusting, and measuring progress toward organizational goals and priorities factors heavily into the discussion over the purpose of performance evaluation systems in the literature. Authors suggest that performance evaluations may even instigate the process of "setting goals or objectives" for organizations (Hatry, et al. 1990) or to even define what good performance is in an organizational context (Kopczynski and Lombardo 1999). In contrast, once good performance is recognized and distinguished, organizations may



also attempt to use performance evaluations to “introduce consequences for performance” that is unsatisfactory, an interesting concept that has animated much vigorous debate especially when the results of performance appraisals are directly linked to employee compensation or the possibility for termination (Osborne and Plastrik 2000). Adjustments to organization goals or priorities may also be made on the basis of such performance evaluations, authors have found, noting that some organizations use them in “modifying program plans to enhance performance” (Hatry, et al. 1990). One of the most commonly identified purposes involves organizations using these programs in “monitoring and evaluating the results to determine if they are making progress in achieving the established goals and objectives” (Hatry, et al. 1990, Wholey and Newcomer 1997, Osborne and Plastrik 2000, Kravchuk and Schack 1996).

As assessments are made of organizational, division or individual effectiveness, performance evaluations also aid in identifying strengths and weaknesses and where to improve or take action to better perform against stated objectives, authors argue (Wholey and Newcomer 1997, Hatry, et al. 1990, Kopczynski and Lombardo 1999, Osborne and Plastrik 2000, Kravchuk and Schack 1996). This process of organizational goal-setting and the use of performance evaluation programs as fundamental managerial tools which refine and assist with this process is perhaps one of the most widely-referenced purposes which authors suggest as motivating the use of performance assessment measures. Since an organization’s ability to define and efficiently pursue its goals is seen as critical to its effectiveness and success in the context of the private sector-oriented excellence literature which has been previously discussed, this emphasis on using performance appraisals to assess an organization’s human resources and identify areas for improvement in their pursuit of organizational goals is not surprising. Such literature argued that

“agencies with clearly defined missions and goals are more likely to perform better” and that a tendency toward “goal ambiguity...undermines organizational performance” (Moynihan and Pandey 2005, Peters and Waterman 1982, Rainey and Steinbauer 1999, Wilson 1989, Meier 2000, Rainey, et al. 1995). As Moynihan and Pandey explain, “A main goal of the NPM [New Public Management] and reinvention-style reforms was to introduce a strategic ‘clarity of task and purpose’ to public organizations through a variety of organizational reforms (Moynihan and Pandey 2005, citing Holmes and Shand 1995).

Robert Behn attempts to condense all these sources into one comprehensive understanding of the different purposes for which performance evaluation systems might be created. He provides a list of eight categories of purposes that may motivate managers to measure performance. Again, Behn’s analysis of the purposes behind performance measurement systems is contextualized to public sector systems. He states:

From the diversity of reasons for measuring performance, I think public managers have eight primary purposes that are specific and distinct (or only marginally overlapping). As part of their overall management strategy, the leaders of public agencies can use performance measurement to (1) evaluate, (2) control, (3) budget, (4) motivate, (5) promote, (6) celebrate, (7) learn, and (8) improve. (Behn 2003).

These eight categories are further explained below, using the explanation that Behn provides in his analysis (2003).<sup>6</sup> Behn's listing represents perhaps the most succinct reduction of all the discussion and debate in the literature over what purpose performance evaluations exist to serve. Behn even further simplifies the understanding of performance evaluation systems when he states: "For the measurement of performance, the public manager's real purpose- indeed, the only real purpose- is to improve performance. The other seven purposes are simply means for achieving this ultimate purpose" (Behn 2003). While Behn makes sure to note that how many of the other seven "subpurposes" or "distinct means" individual organizations choose to use in pursuit of the overarching end of improving performance "is somewhat arbitrary" (2003). Behn's ultimate point is, however, that there are multiple paths to improving organizational performance (2003). These are represented by the first seven of his eight categories and, in the end, each organization must choose for itself which path to take (Behn 2003). This choice by the organization of which of these subpurposes it believes are essential for enabling it to improve its effectiveness and performance will shape its priorities. Thus, its individual employee performance assessment programs must be designed and implemented in such a way so as to align with its priorities and the mode it has chosen to use to improve performance. In the course of his analysis, Behn emphasizes two fundamental points regarding purpose behind performance evaluations. First, Behn emphasizes the importance and usefulness of performance evaluations

---

<sup>6</sup> Behn expounds upon his eight categories of purposes that lie behind the use of performance evaluations by using questions that the managers might seek an answer to and to which the data provided by the performance assessment would provide an answer. The questions he gives for each of his eight categories are as follows: (1) *Evaluate*- How well is my public agency performing?, (2) *Control*- How can I ensure that my subordinates are doing the right thing?, (3) *Budget*- On what programs, people, or projects should my agency spend the public's money?, (4) *Motivate*- How can I motivate line staff, middle managers, nonprofit and for-profit collaborators, stakeholders, and citizens to do the things necessary to improve performance, (5) *Promote*- How can I convince political superiors, legislators, stakeholders, journalists, and citizens that my agency is doing a good job?, (6) *Celebrate*- What accomplishments are worthy of the important organizational ritual of celebrating success?, (7) *Learn*- Why is what working or not working?, (8) *Improve*- What exactly should who do better to improve performance? (2003).

expressed in conventional wisdom (“what gets measured gets done”) and notes that, because of this, it is essential for organizations to decide upon organizational goals and determine what means they want to use to measure their progress toward these goals (Behn 2003). Second, Behn notes that “there is no one magic performance measure that public managers can use for all of their eight purposes” and stresses that organizations must define what they seek to use performance evaluations to assess and to use a combination of the different metrics and tools available to them to properly measure what they seek to know (Behn 2003). In the end, while organizations may have different ideas about intermediary purposes for having an assessment system, increasing organizational capacity and effectiveness is, plain and simple, the end-all goal of any performance measurement system (Behn 2003). Behn’s overview of the eight purposes for which performance appraisal systems are used, his argument that any performance evaluation system will, as its end objective, aim to improve performance (Behn 2003). Grote’s understanding of performance as comprised of both processes and outcomes (or employee “behavior and results”) lays the foundations for the rest of the literature’s discussion with regard to the design and implementation of performance appraisal systems (Grote 1996). Much of the literature focuses on the concept of manager-employee feedback and its importance as a technique for ensuring that performance appraisal systems serve the purposes for which they were designed. Since feedback is seen by some authors as essential to improving employee performance (although there is some debate over this), some authors are concerned with measuring whether or not formal performance appraisal systems encourage an organizational culture of constant performance-related dialogue between manager and employee. In addition, much attention has long been paid to the subject of whether or not formal performance

evaluation processes produce accurate accounts of employee performance and attempt to explain what may cause the tendency for performance reviews to provide skewed or inaccurate information.

### The Importance of Implementation, Communication and Creating a Culture of Feedback.

The issue of feedback is discussed frequently throughout the literature and several authors highlight its importance to the success of any performance appraisal system. According to a study conducted in 2004 by Hinkin and Schriesheim, survey results confirmed “well-established behavioral psychology theories” and suggested that feedback and constructive communication from managers to employees had significant effects on “employee performance” (cited by Marlinga 2006). The ultimate conclusion of the study was that “employees show higher levels of job satisfaction and improved performance when they receive more feedback” and that “omission of any feedback at all is extremely detrimental to employee satisfaction” and effective employee performance (Marlinga 2006, citing Hinkin and Schriesheim 2004). Other studies have corroborated these findings, including one of 851 paper industry managers who “were asked to rank items they felt would lead to better performance” on the part of their employees (Marlinga 2006, citing Beary 1990). With “over thirty items to choose from including involvement, pay for performance, recognition and training,” the item that managers selected with the most frequency was “open and honest job performance feedback” (Marlinga 2006, citing Beary 1990). Using studies similar to those by Beary (1990) and also by Hinkin and Schriesheim (2004), authors have generally accepted the notion that

those performance evaluation systems which facilitate frequent communication and which give managers ample opportunities to provide constructive feedback to employees are the systems which will be most effective at improving employee performance. A primary contention in the literature has been that employees, including those in public sector organizations, seek honest appraisals of their work and suggestions of how to improve and are dissatisfied in working environments with performance assessment systems which do not encourage a culture of feedback and engagement between levels of management and their employees. Of public sector employees at all levels, write Kee and Black, “virtually all want to do a good job” (1985). They note that: “The ‘excellence literature’ observes that people, without regard for where they work, need to succeed, to be winners, to feel that they contribute... It is clear that people today are looking for more than a paycheck from their nine-to-five job... They are concerned with financial benefits, but they want work to have meaning” (Kee and Black 1985). They argue that this tendency for employees to seek satisfaction and deeper meaning within their work means that employees seek communication about their value to the organization and clearer communication about organizational expectations of them so that they might have the satisfaction and motivation derived from improving their performance to meet organizational objectives and to fulfill the expectations they have of themselves.

Most of the literature agrees with the crucial importance attached to the issue of feedback as represented by these studies. Feedback as one of the most fundamental elements in the actual improvement of employee performance (Longenecker and Ludwig 1990, Lee 1985, Deets and Tyler 1986, Ghorpade, et al. 1995). As Behn argues: “Improvement requires attention to the feedback...Improvement is active, operational learning” (2003). Employees have often been

observed to react negatively, become increasingly dissatisfied with their work and, as a result, perform poorly, when there is “an absence of continuous feedback” (Singh, et al. 1981).

Significantly, the use of performance evaluation systems has been suggested as essential to producing increased levels of manager-employee feedback (Sayeed and Bhide 2003, Taylor and Zawacki 1978). Since performance evaluation systems are intended to measure and, ultimately, to improve performance and since feedback is seen as one of, if not the most important element in improving performance, the importance authors attach to feedback in the design of performance appraisals is not surprising. For Behn, “Performance ‘measurement’ is not an end in itself, but may be used by managers to make improvements” (2003). Since the increased levels of feedback engendered by performance appraisal systems are an effective means to improving individual and organizational performance as some authors assert, facilitating an environment in which such feedback can occur becomes the purpose behind the use of such performance appraisal systems for these authors.

Increased flow of communication and feedback from managers to employees is seen as at least one of the intended results of or purposes behind having such systems in the first place (Longenecker and Ludwig 1990, Lee 1985, Singh, et al. 1981, Spriegel and Mumma 1961, Henry 1962, National Industrial Conference Board 1968, Barrett 1966, Williams, et al. 1977, DeCotiis and Petit 1978). Other authors see feedback as less important, merely enhancing or adding to the overall performance appraisal process (Lee 1985, DeCotiis and Petit 1978). Still others identify it as a “side benefit” of a well-designed performance appraisal system (Pareek 1976). Under Behn’s framework with his eight purposes for which performance appraisals are used, he notes that the importance of feedback to appraisals falls in line with his fourth purpose

(motivation of employees) and, more importantly, his eighth (improvement of overall performance) (2003). In some studies, the authors have essentially judged the effectiveness of performance appraisal systems in part by how much increased feedback employees received as a result of their use (Sayeed and Bhide 2003). To this end, authors point out that the more performance assessments aim at increasing the flow of authentic and honest feedback between managers and employees the greater the likelihood that the performance data that is produced is accurate and thus the greater the chance that performance evaluations serve their intended purpose (Lee 1985, DeCotiis and Petit 1978). Authors have also argued that refocusing performance evaluation efforts on ensuring the production of meaningful feedback for employees has often been seen as an essential element in improving faulty or dysfunctional performance assessment programs (Deets and Tyler 1986).

It has been argued that performance appraisal systems in which feedback is an important component more effectively meet employees' desires for honest assessments of their performance in order to help themselves improve and, as a result, feel more motivated and satisfied with their work. This is itself a significant factor in improving performance and is perhaps why feedback is so essential to the performance management process (Ghorpade, et al. 1995, Locke and Latham 1984, Pritchard, et al. 1988 Behn 2003). The authors often cite feedback's ability to strengthen employee "self-efficacy" or performance by edifying employee perceptions of themselves and their capabilities as the reason for feedback's essential role in a performance appraisal system (Campbell and Lee 1988, Behn 2003). In other words, because feedback is motivating and helpful in improving job satisfaction for employees, it necessarily improves performance and therefore ought to be a primary component of any effective system



designed to evaluate and improve employee performance. Many authors qualify their discussion of the issue of feedback, arguing that, in order for feedback to truly inspire and motivate improved employee performance, it must be behavior-related (Lee 1985, Campbell and Lee 1988, Carroll, et al. 1985, Herold and Greller 1977). For this reason, some authors specifically suggest the use of “feedback charts” and other such instruments in the performance evaluation process, encouraging employees to constantly and informally document and assess their behaviors against frequently-communicated manager expectations (Campbell and Lee 1988, Locke and Latham 1985, Bernardin 1981, Dunnette and Borman 1979). Self-documentation and reflection of employees on their behavior relies on and is thought to be inspired by accurate, honest and constructive managerial communication (Campbell and Lee 1988).

Performance appraisal systems which do not encourage or cultivate an organizational culture of constant feedback and increased communication between managers and employees are often seen as having failed to bring about what is considered a core component of their success in turning around poor performance results or instigating and sustaining excellent performance (Longenecker and Ludwig 1990, Lee 1985). Such constant dialogue, often termed a “feedback loop” has been deemed crucial to an effective system (Campbell and Lee 1988, NAPA 1994, Behn 2003). As the National Academy of Public Administration states: “Ideally, performance data should be part of a continuous feedback loop that is used to report on program value and accomplishment and identify areas where performance is weak so that steps can be taken to promote improvements” (NAPA 1994). Sayeed and Bhide agree: “If [a performance appraisal system] is designed and implemented well, it would surely create better team work on an ongoing basis due to established feedback channel among superiors and subordinates that has

been fostered with conscious efforts by the top management support” (2003). As a result, authors often suggest that regular, consistent feedback, even outside of formal performance reviews ought to become normalized in order to truly improve employee performance (Weiss 2004, Lee 1985). As Longenecker and Ludwig advise when discussing the components of an effective performance assessment program, “Managers must provide ongoing feedback, making performance appraisal an ongoing process and not a once or twice per year bureaucratic event” (1990). Because of feedback’s importance in the success of any performance evaluation program, many authors discuss how essential it is for managers to be trained or coached on how to properly and effectively deliver meaningful feedback that identifies areas for improvement without being overly evaluative so as to be demotivating (Lee 1985, Bernardin and Beatty 1984, Singh, et al. 1981). Other authors suggest that, when performance evaluations are used to encourage greater levels of feedback for employees, managers should be encouraged to “be non-evaluative, or non-judgmental, when giving feedback” in the context of a formal performance review (Blanchard 1996). Some authors have also suggested that performance-related feedback ought to come from clients or customers and co-workers or peers in addition to an employee’s supervisors to employees (Poister 2010).

Some authors, however, while still seeing feedback as an essential component of effective organizational and individual performance, do not view performance appraisal systems as being strongly linked to greater employee development or an increase in the level of feedback received. Some go so far as to argue that performance appraisal systems are certainly *not* essential or necessary for bringing about the effective managerial practice of giving feedback about performance to their employees (Beary 1990, Weiss 2004, Blanchard 1996). As Beary states: “A

professionally developed, state-of-the-art performance evaluation program is not the answer to feedback problems...A person can get good performance feedback where no formal program exists and and poor feedback where an excellent program exists (Beary 1990). These authors sometimes express doubt that performance evaluation systems will bring about such essential performance-related dialogue in the workplace, saying that those performance assessments which actually result in the ideal of valuable, consistently-delivered feedback are rare (Campbell and Lee 1988). Such authors stress that feedback is part of a larger phenomenon that must be encouraged outside of the formal appraisal system (Weiss 2004). They argue that effective organizations are marked by continuous feedback that exists outside of structure, organized performance review meetings and is instead the result of a strong, open and honest working relationship between manager and employee (Weiss 2004). Others caution that performance appraisal systems, as they have been traditionally constructed, may do little to reduce the ambiguity that often surrounds manager expectations of employee performance due to lack of communication and that a reliance on such systems to serve this purpose may be unfounded (Campbell and Lee 1988, House and Rizzo 1972, Jackson and Schuler 1985, Rizzo, et al. 1970).

### Are Performance Evaluations Useful and Under What Circumstances?

The literature is somewhat divided, as might be expected, on whether or not performance evaluations of employees are valuable tools for improving organizational effectiveness. Some authors contend that performance evaluations are not worth the effort and time spent on them; they argue either that they cost more in time and resources than they produce in benefits, that

they actually hamper organizational performance and effectiveness or that their results are inaccurate and thus not worthy of being gathered in the first place. Others argue that performance evaluations are invaluable tools for managers in the deployment of resources and the leadership of effective teams and groups toward accomplishing organizational goals. They claim that such measurements assist in the effective deployment of resources where they are most needed and that they allow for an organization to improve its human resources by helping individual employees to maximize their potential and improve their performance through feedback. Most of the disagreement centers around the consistently-reported problem of accuracy in the ratings produced by such systems, especially when the ratings are made by an employee's direct supervisor or manager. Some of the disagreement concerns the effects- the advantages and disadvantages- on overall organizational effectiveness that such evaluation systems have.

Discussions of the effect performance evaluations have on employee motivation, morale and, ultimately, as some authors argue, performance, factor heavily into the debate over the usefulness of formal performance appraisals. In some cases, authors argue that by punishing and demoralizing employees for poor performance instead of helping them to improve, evaluations hamper overall performance. In other assessments of the usefulness of such evaluations, authors see performance evaluations as significant (and sometimes as the most important) tool available to managers *and* employees in improving performance organically through constructive feedback. Some authors prefer to examine the employee-supervisor relationship and how this is affected by the use of performance evaluations. Others begin by attempting to determine whether even the idea of assessing employee performance is morally justifiable. Occasionally there are discussions of performance evaluations that identify the problems with such systems as lying in

the design or implementation or the lack of clear communication about manager expectations of employees during the evaluation process. Two of the major challenges that have been identified by those attempting to design performance appraisal systems lie with both the the systems' design and also its implementation. One set of problems arises in attempting to fashion an evaluation method which will accurately measure performance, something which has been more difficult for researchers to create than it might seem; the other arises from the problem of how to "fairly" judge whether or not performance is "excellent" once it is accurately measured (Taylor, et al. 1995, Kee and Black 1985). As some scholars have identified, "In many organizations performance appraisal systems remain one of the great paradoxes of effective human resource management. On the one hand, appraisal systems can provide valuable performance information to a number of critical human resource activities...On the other hand, there is evidence that appraisal systems are a practical challenge to the academics who often design them and to the managers and employees who must use them" (Taylor, et al. 1995).

Some of the most outspoken critics of performance appraisal systems have often called for an end to the long-lasting trend of using them in private sector corporate, and now public and governmental, settings (Davis and Landa 1999, Heathfield 2000, Nickols 2000). Some have expressed reservations about the usefulness of performance appraisals, noting that "a gap exists between the intended and the actual results" (Padovani, et al. 2010). Other scholars have begun to "question the value of performance measures" and to identify a discrepancy between "rhetoric and reality" (Padovani, et al. 2010, citing Bouckaert and Peters 2002, Chan and Gao 2009, Midwinter 2008, Streib and Poister 1999, Taylor 2009). In one particularly negative review, the author notes that performance evaluation "devours staggering amounts of time and energy, it

depresses and demotivates people, it destroys trust and teamwork and, adding insult to injury, it delivers little demonstrable value at great cost” (Nickols 2000).<sup>7</sup> In his argument against the use of performance appraisals, Nickols asserts that there are fundamental problems with performance evaluations which make them worthless to the organizations that use them (2000). Nickols contends that such systems often produce inaccurate results, are based on unfair and subjective criteria, are demoralizing and demotivating exercises for employees, are harmful for organizational performance and, as a result of these disadvantages to organizational output and productivity, are not worth their enormous cost (2000). Such views and the research off of which they are based, while extreme, have significantly influenced practitioners’ use and perception performance evaluations both in the public and private sectors. One report published in *Personnel Today*, a periodical devoted to discussion of current trends in human resources practices, noted from a survey that the view has gained traction in the human resources industry, saying, “two-thirds of human resource professionals ‘have little or no confidence in their organizations’ performance appraisal systems” (Marlinga 2006, citing *Personnel Today* 2005).

---

<sup>7</sup> Fred Nickols’ article entitled “Don’t Redesign Your Company’s Performance Appraisal System: Scrap It!” appeared in the *Corporate University Review* in June 2000. In the article, he listed 11 major reasons why performance appraisal systems were not worth pursuing. These were as follows:

- 1) The reductions in productivity caused by formal evaluation systems
- 2) The erosion of actual effective performance on the part of employees that was brought about
- 3) The tendency of systems to create emotional anguish for employees
- 4) The evidence that performance evaluations are damaging to employee morale and motivation
- 5) The short-term view that is fostered by the use of these types of systems
- 6) The harm done due to the misguided emphasis placed on the individual, as opposed to the team, and the task, as opposed to the process, by these systems
- 7) The institutionalization of existing values and biases and opposition to change fostered by these systems
- 8) The fear and lack of trust that are fostered by these systems
- 9) The carrot-and-stick management system which develops as a result of these systems
- 10) The impossibility of redesigning formal performance appraisal systems to adjust for these problems
- 11) The excessive cost of system design, implementation and maintenance, which is not worth it, given all the disadvantages such systems pose for organizational success.

Some authors who decry the problems with performance evaluations focus on issues caused by the items on which ratings of performance are based, arguing that, as usually the sole or most significant component of employee evaluations, supervisor ratings of employee performance are inherently subjective and biased and create an unfair standard by which to define excellent performance and against which to rate the performance of individual employees (Marlinga 2006, Nickols 2000). These authors argue that the only way “a system can be effective is if the manager’s perceptions are objective, accurate and completely free of any bias” and that this rarely, if ever, is the case (Marlinga 2006). Many authors document the frequently discussed issue of rater inaccuracy and the question of why managers might be motivated to give inaccurate depictions of employee performance (Longenecker and Ludwig 1990, Lee 1985, Singh, et al. 1981, DeCotiis 1978, Gellerman 1976, Davis 1974, Cherrington and Cherrington 1974, Taft 1971, Kay, et al. 1965). This consumes much of the discussion of performance evaluations in the literature and it is noted that this has been a long-standing problem since the first performance appraisal systems began to be used (Longenecker and Ludwig 1990, Lee 1985). Several authors have cast doubt on whether or not performance evaluations can be held to be ethically or morally justifiable at all in light of the fact that they inherently involve “judging another human being using a rather subjective process” (Longenecker and Ludwig 1990, Banner and Cooke 1984). These authors often support the contention that whether or not a particular performance evaluation can be considered as fair is necessarily contingent upon the very practical issues of the system’s design and implementation, such as the standards against which managers are asked to rate employees (Longenecker and Ludwig 1990, Longenecker, et al. 1988).

These authors argue that, in order to be legitimate, such systems ought to be implemented with effort to provide an “even application” of the standards across the board and with a concern for fairness throughout the process in an attempt to provide accurate reporting and depiction of employee performance against clearly communicated and attainable performance goals (Longenecker and Ludwig 1990). Authors accuse systems of being unfair most often when, as a result, “individuals are punished or held accountable for things that are out of their control and which are really ‘natural variations in system or process performance’” (Marlinga 2006, citing Nickols 2000). Some authors have estimated that ninety percent of all supposed problems with performance identified by performance evaluations are attributable to these natural, systemic or “common causes” of fluctuation in performance (Ghorpade, et al. 1995, citing Deming 1986). In the opinion of these authors, blaming employees for these performance issues, which are arguably not within their sphere of influence, is patently unfair (Ghorpade, et al. 1995). They state, as a result, “Current performance appraisal practices are unfair since they hold the worker responsible for errors that may be the result of faults within the system” (Deming 1986).

One way in which authors question the fairness of performance evaluations is in drawing attention to whether or not the standards against which employees are judged are themselves able to genuinely capture and define what the employee’s ideal performance ought to be (Singh, et al. 1981). This issue has been frequently raised by authors (Sayeed and Bhide 2003, Ghorpade, et al. 1995, Campbell and Lee 1988, Lee 1985, Singh, et al. 1981). Some authors identify unintentional, but significant flaws in an organization’s development and use of certain performance data above others in the evaluation of an employee (Lee 1985). Sometimes this takes the form of managers or supervisors either not knowing what constitutes effective



performance in the case of a particular employee or not having the ability to properly assess that employee's performance against the standards given (Lee 1985). In other situations, managers may know what the standards are and have the resources to observe and assess employees, but fail to communicate these to the employees and thus employees are unaware of what is being used to evaluate their work (Ghorpade, et al. 1995). Finally, even when both parties involved have knowledge of the standards, there is still the question of whether or not the standards are themselves a fair and accurate description of what an employee's performance ought to be (Ghorpade, et al. 1995, Campbell and Lee 1988). As Lee notes: "As a result, raters often form their own reality on the basis of information available to them, selectively attending to some behaviors while ignoring others. Performance rating thus may be based on relevant and irrelevant criteria or behaviors unlikely to be representative of the ratee's job-related behaviors" (1985, citing Bernardin and Beatty 1984, Ilgen and Feldman 1983, Feldman 1981, and Hogarth 1980). This in itself is unfair since the performance of employees is being judged and rated against indicators which may not even adequately capture what the ideal performance expectations of an employee ought to be.

#### The Issue of Accuracy in Performance Evaluations.

These authors insist that the determination of the usefulness of a system to an organization has to be grounded, for ethical and moral reasons, in the accuracy of its reporting and the clarity and objectiveness of its standards (Longenecker and Ludwig 1990, Longenecker, et al. 1987). When these components are missing, the system must be judged as unfair to the

evaluated employees and thus ineffective. On the basis of the assertion that any system which is useful to an organization must be fair (i.e., it must produce an accurate depiction of actual employee performance), and also considering that having a performance appraisal system which consistently produces such accurate data is a rarity, some authors have concluded that performance appraisals are not useful tools for organizational management. Longenecker and Ludwig acknowledge that, on the basis of these arguments, “we would be hard pressed to say that organizations are morally justified in their continued use of performance appraisal systems” (1990). An unfair system is of little significant worth to an organization for practical, decision-making concerns, since “human resource decisions are based on the assumption of accuracy” and such decisions, if based on unclear standards or even if performance against clear standards is inaccurately reported, are inefficient and detrimental to an organization’s attempt to achieve its performance goals (Longenecker and Ludwig 1990). In addition, the authors assert that, in these situations, performance evaluations systems are also useless to an organization because of the simple fact that they are unfair to employees. On these grounds alone, the authors argue, such systems should not be permitted to continue in operation (Longenecker and Ludwig 1990).

Authors have also been quick to criticize performance evaluation systems as not being useful for firms because they result in weak and inaccurate knowledge about the nature of performance itself (Marlinga 2006, Nickols 2000, Ghorpade et al. 1995, Deming 1986). Some authors argue that, due to the inherent set-up of formal performance evaluations, especially when they are tied directly to organizational rewards or punitive actions such as compensation, bonuses, tenure, or other such consequences, the evaluation data recorded is actually skewed due

to employees or managers attempting to inflate or improve how performance reflects on the evaluation, as opposed to how it actually occurred (Nickols 2000, Ghorpade, et al. 1995, Longenecker and Ludwig 1990, Lee 1985, Singh, et al. 1981). This entire issue has already been discussed under the section on the issue of accuracy in performance appraisal rating earlier in this study. When performance evaluations are closely tied to benefits, employees, in addition to managers, have an incentive to inflate either the actual ratings or the temporary appearance of their performance during observations to those who are tasked with giving the ratings. In these systems, authors argue, “managers who write the appraisals [become] the targets of efforts aimed at influencing and manipulating” performance ratings and “‘brown nosing’ on behalf of the employees” is indirectly encouraged (Nickols 2000). This in itself results in an inauthentic portrayal of employee performance and a misuse of organizational resources. Sometimes, as a result, employees inflate evaluation data intended to reflect their performance by setting overly easy or simple goals for themselves in order to boost their performance scores (Nickols 2000). The measurements produced by these performance reviews then assess employee accomplishments at tasks they design and incorporate into their performance evaluation, but which do not actually represent effective measures of their true performance (Nickols 2000). This occurs most often in the systems which utilize customized goals and employee progress against those goals as the criteria of the evaluation.

### The Issue of Employee Morale and Motivation.

Employee scores also have the potential to become deflated for a number of reasons and, when this happens, employees can become demoralized or can lose motivation. As a result, productivity may lag as employees lose trust in the system and feel that evaluation results or criteria and standards are arbitrary or overly subjective. As Ghorpade, et al. note: “Current performance appraisal practices create a band of discouraged workers who cease trying to excel” (1995). Other authors have noted that employees have the tendency to view “formal performance appraisal sessions [as] demoralizing and anxiety-producing encounters” and that this has disastrous effects on productivity (Campbell and Lee 1988). It is important to understand that employee demotivation has significant consequences for organizational productivity which is why authors are increasingly urging that these side effects of performance appraisals be taken into account as they are being designed and implemented (Ghorpade, et al. 1995). Systems which use a forced distribution or forced rankings model are particularly criticized by researchers for this problem. As some have noted on the basis of a survey of human resources professionals conducted on the subject, “forty-three percent of [human resources professionals] said forced rankings reduce collaboration and teamwork and one-third stated forced rankings resulted in higher turnover costs” (Marlinga 2006). Researchers note an increase in feelings of frustration and dissatisfaction when employees feel that organizational rewards or punitive measures are unfairly or arbitrarily distributed on the basis of inaccurate or unfair standards or performance data (Davis and Landa 1999, Grote 1996). As a result of these types of systems or others in which employees feel as though the performance data produced was poorly produced, employees

may feel that there is no reasonable justification for why they are rated in the way they have been by their supervisors compared to other employees and, as a result, feel increasingly dissatisfied and become less productive.

Additionally, researchers have stated that employees increasingly become dissatisfied and demotivated when they feel as though they have been rated against objectives or standards which were not clearly communicated to them in the first place (Singh, et al. 1981, Labovitz 1972). This particular issue has been confirmed in survey research and, it is noted, usually occurs, as might be expected, when employees receive negative ratings and feel as though these ratings are unjustified (Glickman 1955). Other authors have argued that, due to the format of most performance evaluations which require the rating of a manager or supervisor, “obtaining accurate appraisals of the employee’s job behavior” was nearly impossible (Smith and Kendall 1963). In an article published in *Personnel Today* in July 2005, roughly “two-thirds of human resource professionals ‘have little or no confidence in their organizations’ performance appraisal system” (Marlinga 2006). Employee demoralization remains a significant reason behind why researchers often accuse performance evaluations of resulting in decreased productivity and thus declare them not worth the effort (Marlinga 2006). As some argue: “Apart from those few employees who receive the highest possible ratings, performance review interviews, as a rule, are seriously deflating to the employee’s sense of worth. Not only is the conventional performance review failing to make a positive contribution, but in many executives’ opinions it can do irreparable harm” (Davis and Landa 1999). As others have noted, “particularly when appraisal systems are viewed as unfair, the effects on morale are devastating” and this ultimately has consequences for employee performance (Marlinga 2006).

In studies on the effect of performance appraisals on employee motivation, research has confirmed that, even before an employee sees the results of his or her performance evaluation, he or she might begin to put up “defense mechanisms” out of fear of “lowering favorable self-perceptions” and in order to “preserve a positive self-image” even if such an image is, according to the evidence and the performance appraisal, inaccurate (Campbell and Lee 1988). This can result in frustrations with managers and in a lack of personnel development for those employees who need it most, according to the results of the evaluation (Longenecker and Ludwig 1990, Ghorpade, et al. 1995). Ultimately, employee productivity wanes as employees, due to the pervasive fear of formal performance evaluations, shut themselves off from meaningful, constructive criticism designed to improve their performance and fashion for themselves an inaccurate, positive image of their own performance (Campbell and Lee 1988). In this way, the potential for improved individual and organizational performance is lost. In such a situation, employees are less likely to seek help and less likely to improve performance. In this case, as productivity declines, it is clear that performance evaluations are failing to bring about their most fundamental intended result- improved performance. In some cases, the decline in productivity can actually be quite severe. As some have noted, “it’s estimated that in some cases, reduced productivity following a review may last from three to six months, resulting in real costs to the organization” (Marlinga 2006).

### Issues Involving Communication and Implementation in Performance Evaluation Systems.

Poor communication and ineffective system design and implementation have also drawn the attention of researchers in this field; they also play a significant role in these instances of employee dissatisfaction and, more significantly, in the reasoning of some of the authors as to why performance evaluations are not worth utilizing (Marlinga 2006, Nickols 2000, Davis and Landa 1999). Certainly the difficulty involved in successfully designing an effective performance evaluation is important, however, increased attention has been paid to how many problems arise in system implementation and administration and many have raised doubts about whether or not these challenges are worth confronting (Nickols 2000). While the research on performance evaluations has tended, as some authors point out, to focus more on design and creation of the ideal performance evaluation system, the problems that poor implementation poses for the effectiveness of the system overall are being increasingly addressed by more researchers (Longenecker and Ludwig 1990). Yet, as these authors note, sometimes the most significant of a system's flaws can occur during the process of its implementation (Longenecker and Ludwig 1990). A 1997 study involving a Canadian consulting firm's performance evaluation system is cited by some of the authors as an indicator of how poor system implementation can have disconcerting results for system effectiveness, overall organizational success and individual employee performance (Davis and Landa 1999). In the study "only about two-fifths of the sample (42%) reported regular, timely performance reviews" and "less than half (47%) said that their managers clearly expressed goals and assignments" (Davis and Landa 1999). Because of

this, “less than two-thirds of the sampled employees (60%) said that they understood the measures used to evaluate their performance” (Davis and Landa 1999).

The link between such a poor implementation and understanding of goals and objectives and the effects this has on performance is made clear. Initially, what resulted were increased feelings of employee demoralization and dissatisfaction as noted above. As a result, few “sampled employees (57%) thought that their performance was rated fairly” and very few (19%) reported “a clear, direct, and compelling linkage between their performance and their pay” when such a connection was supposed to be made (Davis and Landa 1999). Because of this, employees also felt as though their performance had not been improved through the experience. Only 39% “reported that their performance review was helpful in improving their on-the-job performance” (Davis and Landa 1999). Similar studies have confirmed these results including one which analyzed the results of a survey completed by human resources professionals from around the country. The survey “noted that two-thirds of human resources professionals ‘have little or no confidence in their organizations’ performance appraisal systems’” and “40 percent said the process did not achieve what it set out to do” (Marlinga 2006). Large numbers of respondents to this survey also felt that “performance appraisal systems had a negative impact on an individual” and that “appraisals were often badly conducted” (Marlinga 2006).

Communication is essential for an effective manager-employee relationship, for effectively improving performance and, ultimately for the success of any firm attempting to use such an evaluation system. Many authors argue, however, that performance appraisals are so frequently plagued by a lack of communication between managers and employees and that this represents such a significant issue that it necessarily calls into question the effectiveness of any



system (Ghorpade, et al. 1995). Many authors confirm that, in typical performance evaluations systems, “employees often do not have a clear idea of what their supervisors expect” and this poses significant problems for communication and for overall individual and organizational effectiveness (Campbell and Lee 1988, citing House and Rizzo 1972, Jackson and Schuler 1985, Rizzo, et al. 1970). Other “research indicates that many employees, even after their evaluation sessions are unaware of the performance criteria being used” (Campbell and Lee 1988, citing Brief, et al. 1981). Sometimes, even when managers try to communicate effectively, the result is only “minimally useful in overcoming ambiguity” because of the format of the performance appraisal system or because of the poor communication or conflict resolution skills of the managers (Campbell and Lee 1988, citing Carroll, et al. 1985, Herold and Greller 1977). This demonstrates to the authors that a formal system is useless and not a worthy endeavor for organizations and firms to undertake.

When communication about what expectations for employees are, how employees are performing on a daily basis or even how the evaluation process itself should work is lacking, the effects on performance appraisal can be serious (Singh, et al. 1981). As Singh, et al. note, the confusion that results for employees certainly can lead to a decrease in the system’s effectiveness or value for the organization and ultimately more tension and dissatisfaction for employees, managers and ultimately the organization than would otherwise be the case (1981). They state: “Role ambiguity [which occurs when managers and employees fail to communicate] normally leads to confusion, which consequently affects performance as well as evaluation” (Singh, et al. 1981). Ambiguity about exactly what is expected from an employee plays a significant role in producing some of the negative consequences that the authors note organizations experience

after implementing performance reviews. When employees do not know precisely what they are supposed to be working toward or how to accomplish their job objectives, this clearly has a negative effect on their ability to efficiently and effectively perform their work and contribute to the organization's larger efforts. This has prompted some authors to declare that performance evaluations are not useful because they do not achieve their stated objective- improving employee performance and productivity.

An additional issue some researchers have raised with regard to whether or not performance appraisals are useful to organizations involves the attitudinal side effects for organizational culture that often occur as a result. These manifest themselves, some researchers suggest, in the exaggerated emphasis that often gets placed on the individual and their strengths and successes as opposed to a focus on how employees can contribute to the organization and increase organizational potential for success (Nickols 2000). Such an emphasis has been shown to contribute to a decrease in teamwork or collaboration (Marlinga 2006, Nickols 2000, Davis and Landa 1999, Ghorpade, et al. 1995, Deming 1986). Also, researchers have noted that the use of performance evaluations can result in exaggerated importance being placed on "tasks" and results as opposed to the "processes" or behaviors by which employees achieve these results (Nickols 2000, Ghorpade, et al., Deming 1986). As Nickols argues:

The classic performance appraisal system emphasizes individual or task-level performance instead of team or process performance. Appraising individual performance can be a divisive factor in an environment where genuine teamwork is required. Consequently, in times of change, retaining an appraisal system that focuses on individual task performance sends at best a mixed message when

management calls for teams or wants to focus on business process performance instead of individual task performance (2000).

Authors have described these problems in terms of the short-term view that tends to cloud organizational decision-making when these types of attitudes are encouraged and fostered (Marlinga 2006). Authors point out that such a short-sighted, strictly task and individual-focused culture develops usually when organizations tie the results of performance evaluations strongly and directly to organizational rewards and punishments (Ghorpade, et al. 1995, Deming 1986). Often, performance evaluations can “foster a short-term view...since many evaluation systems only focus on the employee contributions within the past year or last six months, contributions over time get overlooked” (Marlinga 2006). The effects of employees being overly focused on their own work and accomplishments within limited segments of time and of not seeing their work as linked to broader team or organizational goals can have devastating effects on an organization’s long-term success and sustainability (Deming 1986). Ghorpade, et al. (1995), citing Deming (1986) attributes these faults to the very nature of performance appraisal systems themselves, arguing that “current performance appraisal practices promote worker behavior that compromises quality.” With the way such systems are set up, these authors argue that “a short-term perspective results” and explains this by saying that, as a result of how formal performance evaluations are usually framed, “energies are directed toward meeting targets and quotas” and that “the attainment of those outputs becomes a central preoccupation of the workers” (Ghorpade, et al. 1995). Longenecker and Ludwig note that performance appraisals can take on “a short-sighted perspective,” and that, when they do, “it will, in the long run, destroy the game” of attempting to improve individual and organizational performance as a whole (1990).

As Susan Heathfield notes, “[Performance appraisal] is incongruent with the values-based, vision-driven, mission-oriented, participative work environments favored by forward-thinking organizations today” (2000).

#### The Manager-Employee Relationship and Its Effects on Performance Evaluations.

Additionally, performance reviews have been accused of fundamentally obstructing and distracting from the working relationship between managers and employees which is, most importantly, the avenue through which feedback, so essential to genuine personnel development and maximized individual capacity, is delivered (Newstrom 1974, Niazi 1976, Marlinga 2006, Nickols 2000). This is primarily identified by researchers as manifest either in the lack of trust that employees increasingly hold toward their managers as a result of performance reviews or the increased fear of managers on the part of employees due to the power managers hold through the evaluation process (Longenecker and Ludwig 1990). Both of these are unnatural and contrary to the way managers and employees ought to work and operate together. They essentially make the existence of an effective working relationship between managers and employees impossible. Some authors even go so far as to describe the evaluation process as “a highly threatening experience” for employees, or at least that it can be perceived this way (Campbell and Lee 1988, citing Meyer, et al. 1965, Kay, et al. 1965). When faced with these situations, employees can often become defensive and cease to be receptive to any feedback or constructive information which their manager attempts to share with them.

Even when the information being shared is solely for their benefit and development, employees' reactions can still be negative due to their instinctive reaction to formal performance appraisals in general (Campbell and Lee 1988, citing Meyer, et al. 1965, Kay, et al. 1965). The problem of performance evaluation systems fostering a lack of trust or fear of management on the part of employees and its decidedly negative effects on overall organizational success is noted by many authors in the appraisal-related literature (Marlinga 2006, Nickols 2000, Davis and Landa 1999, Longenecker and Ludwig 1990). Without the notion that "trust is always the basis of the manager-subordinate relationship," organizations cannot thrive, researchers argue (Longenecker and Ludwig 1990). Heathfield concurs in this assessment, claiming that the framework of traditional "performance appraisal...smacks of an old fashioned, paternalistic, top down, autocratic mode of management which treats employees as possessions of the company" and goes against what ought to be a very natural and productive employee-manager working relationship (2000). Nickols suggests that "traditional performance evaluation systems foster fear and lack of trust due to a 'carrot-and-stick' management approach" (2000). He notes that there is a "degree of fear associated with the appraisal system" (Nickols 2000). He describes this, stating: "This ties to a lack of trust in one's boss, and management in general, and leads to a phenomenon known as 'malicious compliance,' that is, a passive-aggressive stance of 'tell me what you want me to do and I'll do it' on the part of the employee" (Nickols 2000).

In surveys, where employees have noted that they disagree with the ratings their managers have given them and when they see performance evaluations as "a figurative whip in the hands of management," they are also more likely to fail to see improvements in their own performance or be satisfied with the work they do (Davis and Landa 1999). Heathfield describes

the difficult positions in which performance evaluations place both employees and managers, saying, “The manager must act as judge and jury and be ready to defend any ratings below outstanding. The employee then becomes the anxiety-ridden ‘defendant.’ Both employer and employee become defensive and tense. Since most managers strive to create a harmonious work environment, they are uncomfortable with the level of conflict that performance appraisal systems foster” (2000). The results of this for organizational productivity are clear. Researchers have argued that “reduced productivity and mistrust of management” are inseparably linked and argue that performance evaluations can result in decreased “collaboration and teamwork” which can have extremely detrimental effects on organizational culture (Marlinga 2006). Such negative effects for the employee-manager relationship are counter-productive to an organization’s effectiveness and are a serious disadvantage of performance evaluations, some researchers argue. For them, this constitutes one of the major reasons for why formal performance appraisals are not worth an organization’s time and effort and ought to be discarded (Marlinga 2006, Nickols 2000).

#### The Cost of Maintaining and Implementing a Performance Evaluation System.

Finally, in assessing the arguments against the usefulness of performance evaluations, the issue of whether the cost imposed by the difficult task of designing, implementing and maintaining these systems is worth it has been raised by numerous authors cannot be ignored. According to Nickols, “the actual time and financial costs incurred for performance appraisal systems are excessive” (2000). Nickols also notes that performance appraisals have the potential

to draw time and resources away from the actual work of the organization (2000). He compares the costs and benefits of performance appraisals by declaring: “Specifically, the time and money spent preparing, writing, reviewing, copying, filing, distributing, and conducting appraisals in addition to the time training staff regarding the performance appraisal system and in defending post-appraisal appeals, grievances and lawsuits is not worth any possible benefits” (Marlinga 2006, citing Nickols 2000). In light of his assessment of the strong organizational disadvantages and ineffectiveness of performance appraisals, Nickols concludes that such extensive costs are not worth undertaking (2000). He argues: “A reasonable person would be hard pressed to argue that the benefits of performance appraisal systems outweigh their costs. The costs are extraordinary and many of the supposed benefits cited do not withstand serious scrutiny” (Nickols 2000). Longenecker and Ludwig add to the discussion, saying, “We know that organizations spend millions of dollars annually designing and implementing performance appraisal systems that are state of the art” (1990). They also come to the conclusion that, when the disadvantages of performance appraisals are taken into account, it is difficult to justify the use of formal performance evaluations. When such systems produce inaccurate data, which is often used in making significant personnel or compensation decisions and also, on the side, introduce far-reaching negative effects for the overall culture and effectiveness of an organization, they not only fail to serve the purposes for which they have been designed, but actually are more harmful for organizations than if they had not been instituted in the first place, these authors argue (Marlinga 2006, Nickols 2000). When the significant financial and other costs of performance appraisal systems are considered, many authors conclude that such systems are not worth the effort (Marlinga 2006, Heathfield 2000, Nickols 2000, Davis and Landa 1999).

For some authors, when performance appraisals do not serve their intended purpose and cause more harm than they produce in benefits, the resources spent on system design and implementation have necessarily been wasted and are a loss to the organization (Longenecker and Ludwig 1990).

It is important to note that the authors in this body of literature do not completely disregard the argument that improving employee performance is a worthy objective. Nor do they see it as a worthless endeavor for organizations to attempt to improve employee performance. Rather, they see formal performance assessment systems as a harmful, inefficient and costly means to this end and argue that these systems should not be used (Marlinga 2006, Nickols 2000, Beary 1990). These authors are, for the most part, largely supportive of the importance of authentic feedback in improving employee performance and are optimistic that such feedback, when sustainably engendered, can dramatically improve individual and organizational performance (Nickols 2000, Heathfield 2000). They do stress, however, the importance of recognizing that such feedback can occur without the existence of formal performance appraisal (Marlinga 2006, Nickols 2000, Beary 1990). Marlinga concludes that “performance evaluation may not be the most useful vehicle for providing feedback” (2006). As Nickols argues: “Performance-related discussions between bosses and subordinates do not require a formal, full-blown performance appraisal system” (2000). Davis and Landa concur, saying: “It is difficult to understand how this process can be seriously considered a viable method for increasing motivation and productivity” (1999). Instead, “as an alternative, Davis and Landa advocate informal, frequent communication between supervisors and employees in place of formal performance evaluation systems” (Marlinga 2006, citing Davis and Landa 1999). At the very



least, authors who are critical of performance evaluations, such as Brumback, stress “keeping the results of performance appraisals in perspective and [focusing] more on setting performance expectations” (Marlinga 2006, citing Brumback 2003). Even though this report is framed as an analysis of an already-existing performance evaluation system for teachers in Shelby County and in Tennessee more generally, a detailed review of the discussion in the literature on the most significant weaknesses of performance evaluations in general allows for proper emphasis on those components of a teacher evaluation system which might be most fraught with difficulty and which deserve considerable attention.

#### The Significance of Teacher Evaluation Policy to the Education Reform Debate.

As the thirtieth anniversary of the publication of *A Nation at Risk* approaches, many question whether or not the pressing issues brought to light in the “groundbreaking report” by the National Commission on Excellence in Education have been seriously addressed or mitigated (Klein 2011). The report initially confirmed what many analysts had already suspected, “a rising tide of mediocrity that threatens our very future as a Nation and a people,” yet what has actually been accomplished toward combating this trend is difficult to tell (Klein 2011). In his article describing *The Failure of American Schools*, Joel Klein, former Chancellor of New York City’s school system, notes that, to date, “the gains we have made in improving our schools are negligible- even though we have doubled our spending...on K-12 public education” (Klein 2011). Klein goes on to document how, in 2010, only one-third of the eighth graders in the nation were “proficient in math, science or reading,” the high school graduation rate fell below 70 percent

and, according to ACT, “the respected national organization that administers college-admissions tests, recently found that 76 percent of our high-school graduates ‘were not adequately prepared academically for first-year college courses’” (Klein 2011). Klein notes that the result of these failures has a significant effect on the make-up and equality of American society, saying, “The net effect is that we’re rapidly moving toward two Americas- a wealthy elite, and an increasingly large underclass that lacks the skills to succeed” (Klein 2011). Klein argues that the “forces behind reform” are unable to combat a powerful system bent on “defending the status quo” (Klein 2011). Klein suggests that, “without a major realignment of political forces, we won’t get the dramatic improvements our children need” (Klein 2011). Of all the parts of the American education system in need of repair, Klein points to the growing movement to “analyze year-to-year student progress and tie it to individual teacher performance” as the most crucial element in effecting true education reform (Klein 2011).

Since roughly 1980, an increasingly interconnected network of those invested in the national conversation on national education reform has formed and has consistently targeted the development of and implementation of teacher evaluations and other policies aimed at promoting greater teacher effectiveness (Cornett 1995). State legislatures and governors, Congress, the president, the Department of Education, local boards of education and a host of non-profit and for-profit research and advocacy organizations centered around education reform have all played a vital role in contributing to the copious amount of attention paid to education reform in general and teacher evaluations in particular (Cornett 1995). Frequently, these and others invested in the debate over improving educational policy to effect improvements to what is often cited as a poorly-performing system of public education nationwide (Cornett 1995). In these conversations,

the subject of teacher evaluations has been given increased importance and has sometimes been framed as singularly the most important aspect of effective education policy (Koppich 2004, Klein 2011). Of all the individual initiatives that are tied to education reform, the vast majority of the studying, thinking, writing and policy-making that has been done with respect to education has often primarily or solely focused on improving or overhauling teacher evaluation systems (Koppich 2004, Winters 2012). What has resulted has been a “flurry of laws” and an “unprecedented wave of state teacher-evaluation reform across the country” (Cornett 1995, McGuinn 2012).

The movement, to be sure, has clearly articulated its reasoning and the logic behind these approaches. As analyst Julia Koppich argues, “Research has shown what common sense would suggest: Children with high quality teachers are able to make continuous progress. Those with less effective teachers struggle, often not successfully to keep up” (Koppich 2004). Marcus Winters, a senior fellow at the Manhattan Policy Institute, also makes this argument in his report, saying, “Several studies show that to which teacher a student is assigned makes a huge difference in determining how much that child will learn in a given school year. And many of the benefits of good teachers are more long-lasting” (Winters 2012). These arguments have been frequently and thoroughly made and are widely accepted. Nevertheless, this collective effort to transform systems of teacher evaluations “has turned out to be one of the most controversial questions facing reformers” in the entire realm of education policy (Winters 2012). This is because opponents of these reforms do not dispute the validity of the claim that teachers make a crucial difference in a student’s educational experience and academic achievement. Rather, opponents of this increasingly popular and monumental movement raise specific objections to, most notably,

underlying assumptions and ideas about *how* teacher effectiveness ought to be defined and determined, the methods and factors used to assess individual teachers as they measure up against these ideas of effectiveness and whether the emphasis placed on teacher evaluations in the current national conversation on education reform has led to less attention being placed on factors that are equally as important in shaping the success or failure of America's public schools (Rothstein 2010). It is not the importance of teacher evaluations in itself that causes a problem for these critics, but rather the assertion by many advocates of teacher evaluation reform that combating so-called "incompetent teachers" is the "single most important factor" in improving students' educational success (Rothstein 2010). Both sides of the debate invoke concern for students and a desire to see schools succeed, yet take different approaches to achieving these ends when it comes to issues of teacher recruitment, training, evaluation and retention. The focus of this report, as a review of sources, is to provide a brief history of the development of teacher evaluation policy, specifically in the state of Tennessee and in Shelby County and the City of Memphis, to provide a summation of the national discourse and debate on this topic and how it can be brought to bear on local concerns and, finally, to address background factors that give context to and significantly influence national, state and local discussions about teacher effectiveness and reforms to teacher evaluation policy.

#### Effects of Changes to Teacher Licensure and Certification on Evaluation Policy.

Changing teacher evaluation policy must be considered in light of significant changes made to the state law in the late 1980's and early 1990's regarding how prospective teachers

could enter the teaching profession. These laws aimed to address predicted shortages of teachers, there has been an “explosion” of state laws permitting alternative routes for teachers to be certified (National Center for Education Information 2005). In 1983, only eight states had approved alternate routes to teacher certification and licensure (National Center for Education Information 2005). Currently, every state in the nation and the District of Columbia report that they have at least some program for alternative certification of teachers in place (National Association for Alternative Certification 2011). This, according to the National Education Association, “has rapidly evolved into an accepted model for recruiting, training, and certifying those who already have at least a bachelor’s degree and want to be teachers” (2010). As the National Center for Education Information reports, “Not only have more and more states instituted legislation for alternative teacher certification, but also, more and more institutions of higher education have initiated their own alternative programs for the preparation of teachers leading to a license to teach” (National Center for Education Information 2005). This is due to a trend in which policymakers have become increasingly skeptical of traditional licensure programs and schools of education. Increasingly, “the efficacy of traditional systems of teacher preparation has received considerable criticism” (Roth and Swail 2000). For one, these traditional programs were failing to provide teachers for high-need areas, such as school districts in “inner- cities and rural areas,” prompting states to first create “emergency-credential programs” which later became the alternative certification programs of current policymaking (Roth and Swail 2000). Another problem with traditional certification programs where students graduate with education degrees are that they are highly criticized for providing weak preparation for the nation’s future teachers. Advocates of alternative certification programs argue

that we should not be surprised when student achievement fails to reach predicted levels when the students' "teachers are not very good themselves" (Leef 2013). As one commentator argues, "Yes, [traditionally licensed teachers] have their college degrees, but those degrees are easily acquired by some of the weakest students colleges admit" (Leef 2013). In October 2013, education practitioners Barbara Nemko and Harold Kwalwasser authored an article in the Wall Street Journal which criticized the performance of the nation's traditional teacher preparation programs. The authors promoted abandoning the concept of the "education major" and argued that traditional schools of education offer "too much theory" and "not enough practical learning about teaching" (Nemko and Kwalwasser 2013). Nemko and Kwalwasser further argue: "How can new teachers be expected to educate children without first being trained well?" (2013). They state: "The problem, put simply, is that entrance requirements to most colleges of education are too lax, and the requirements for graduation are too low" (Nemko and Kwalwasser 2013). They claim that, in 2010, "the mean critical-reading SAT score of entering college freshmen was 501, but for education majors it was 481. The math score was 516 compared with 486, and in writing, 492 versus 477" (Nemko and Kwalwasser 2013). The result of this has been that any uniformity of standards determining who is permitted to teach in public school classrooms across the country has been virtually eliminated. There are now multiple pathways to teaching for individuals who did not major in a traditional education program in college. Despite this significant change in teacher preparation, certification and licensure policies, however, concern over the quality of teaching and instruction occurring in the classroom has not been reduced. If anything, the number of policies concerned with improving the quality of teachers in public schools has increased.

## What is Teacher Quality?

In a 2011 opinion piece written for the Brookings Institution, economists Michael Greenstone and Adam Looney argue that it is highly important for schools and school districts to prioritize hiring and retaining “highly effective teachers” in an effort to ensure that students are the recipients of what they refer to as “teaching quality” and that they benefit from strong instruction in the classroom (2011). This, however, is not the point of contention for many invested in the debate over education reform. Most involved in the debate over how to reform education in the United States would agree that recruiting, hiring and retaining high quality teachers is an essential element of a well-functioning education system. As the researchers acknowledge, while vastly different policy solutions are proposed to reach this end, the general objective is one with which most people would agree. They state: “Almost everyone, from policymakers to parents to teachers, agrees that reforms in the K-12 educational system are necessary to developing a more educated workforce and a stronger economy” (Greenstone and Looney 2011). As one might expect and as trends in policymaking confirm, the effort to “recruit and retain the best and the brightest in our workforce to become teachers” is the issue perhaps most frequently cited as the “primary challenge” that exists to improving the educational offerings students to which students are exposed, particularly for those from low income communities (Greenstone and Looney 2011).

The reason for this is understandable. Greenstone and Looney note, as have many before them, the strong “economic imperative” for securing “great teachers” and thus bolstering the value of what students take away from their education (2011). Not only does high quality

education “have the ability to transform and enrich the lives and living standards of Americans,” but it is also extremely important in the long run for the larger American economy (Greenstone and Looney 2011). Researchers have identified that, for a class of 20 students, being exposed to high teaching quality as early as the kindergarten level resulted in increased future earnings of more than \$300,000 (Greenstone and Looney 2011, citing Chetty et al. 2010). As Chetty et al. note, “Several other adult outcomes- such as college attendance rates, quality of college attended, home ownership, and 401(k) savings- are also all highly correlated with kindergarten test scores,” which were traced back to the quality of the teacher students were assigned in kindergarten (Chetty et al. 2010).<sup>8</sup> On an aggregate level, Greenstone and Looney argue that a more educated American workforce is on the whole preferable, for a number of economic reasons, to a less educated one (Greenstone and Looney 2011). They state: “A strong educational system and the role of teachers have never been more important for America’s workforce” (Greenstone and Looney 2011). As evidence of this, they argue that “less-skilled workers are disproportionately unemployed and have experienced declining wages” at the same time as “the value of an education in the labor market is at an all time high” with individuals holding a college degree earning roughly “twice as much each year...equivalent to about \$570,000 more [than those without a college degree] over a worker’s lifetime” (Greenstone and Looney 2011).

Most people would agree with these arguments and with the proposition that a strong, effective education system is essential for the wellbeing of individuals as well as for society at

---

<sup>8</sup> To be sure, while the particular focus of this study emphasized the importance of a student’s kindergarten teacher, the implications can be generalized to teachers students experience at each level of their educations. Perhaps the fact that the strength of students’ kindergarten teachers is highly correlated with their success much later in life lends itself to the argument that other teachers who influence students later in life have even more of an impact.



large, particularly when considering economic growth. Most would also agree that highly qualified and effective teachers who provide instruction to students in the classroom are very important components of an effective education system as a whole. These highly generalized assertions are where the agreement stops, however. Considerable debate has occurred, for example, over whether or not a problem with teaching quality exists in the United States and, even if it assumed such a problem exists, over how it should be measured, understood or addressed. In order to comprehensively understand this debate, it is necessary to provide a thorough definition of the term “teacher quality.” At the outset, many involved in the national conversation on teacher quality acknowledge that at least part of the divisive nature of the debate is due to the fact that “teacher quality is extremely difficult to measure” (*Education Week* 2011). For this reason, “the specific characteristics that constitute an effective teacher are hotly debated” (*Education Week* 2011). As education scholar Bruce Torff explains, “disputes arise concerning which teaching skills tend to be the weakest and thus constitute the most urgent threats to teacher quality in our nation’s schools” (Torff 2005). In defining and operationalizing teacher quality, one school of thought has tended to focus on “proxies,” including, but not limited to, “certification, academic degrees, and years of experience” (*Education Week* 2011). This school of thought sees teacher quality as qualifications-based, meaning that the quality of a given teacher ought to be measured through his or her personal qualifications, such as years of experience, level of educational attainment or degree specialization. Teachers with high levels of educational attainment, years of experience, or degrees specializing in the topic they teach in the classroom are considered highly effective or desirable by school districts. For adherents to this

branch of thinking, a problem in teacher quality exists when a given school or district lacks teachers possessing one or a combination of any of these qualifications.

The second main school of thought has attempted to define teacher quality in terms of the effects teachers can have on student achievement, most often measured in the form of student test scores. This school of thought developed in response to the weaknesses identified in drawing concrete links between student achievement and traditional teacher qualifications, such as educational attainment, content-area knowledge and years of experience (Hightower, et al. 2011). This school of thought sees teacher quality as measured through the impact a teacher has on student achievements on standardized tests, intended as the operationalization of successful student learning and academic achievement. Teacher quality is, then, a results-based concept. In other words, highly effective teachers are those who encourage or bring about high levels of learning and strong achievement gains for their students. Teachers who produce these gains are highly effective, while teachers who do not are considered ineffective. There are strong similarities between the two approaches to measuring or defining teacher quality. To begin with, both groups accept the basic premise that effective teachers, however this effectiveness is measured, will, to some degree, produce gains in student achievement, measured by test scores. As many acknowledge, “the preponderance of evidence concludes that effective teachers are capable of inspiring significantly greater learning gains in their students” (*Education Week* 2011). Of the two schools of thought, however, those who argue that a teacher’s effectiveness ought to be measured through his or her student’s test scores place naturally greater emphasis on these scores than do those who attempt to measure teacher quality by assessing more traditional teacher qualifications.

The two schools of thought, as has been noted, share some similarities. They agree on the basic premise that effective teachers will produce results in student learning. Even the school of thought which stresses teacher years of experience, content-area knowledge and certification processes recognize and accept “the preponderance of evidence” which argues that “effective teachers are capable of inspiring significantly greater learning gains in their students when compared with their weaker colleagues” (*Education Week* 2011). In practice, however, these schools of thought disagree on two primary considerations. In addition to disagreeing over how to define and measure teacher quality, both schools of thought also tend to disagree about whether or not student achievement measured in test scores represents the full contribution teachers make to student learning. When results-based researchers see low test scores, they will tend to conclude that the district or school suffers from poor teacher quality, because high quality teachers inspire gains in student achievement, which, they argue are able to be measured in the form of standardized tests (*Education Week* 2011). Qualifications-based authors instead focus on the individual skills, expertise and experience which teachers bring to their school or district. They argue that there are objective, professional standards for determining when teachers are or are not “highly qualified” with regard to a different area of instruction or content (*Education Week* 2011). If a district suffers from low test scores, but has, on average, high quality teachers in each classroom, then these researchers tend to argue that there are other factors affecting student test performance or that test performance is not the best indicator of student learning. They argue that teachers are highly skilled and, therefore, that they are making significant contributions to student learning, but that these gains are not measurable in the form of standardized tests. There

is certainly overlap between the two schools of thought; researchers subscribing to these two schools differ primarily in terms of where they place their focus or emphasis.

Researchers who see teacher quality as inhering in a teacher's individual qualifications or characteristics will be more willing to see these teachers rated as highly effective, perhaps through performance evaluations, even though there may be variations or instability in teachers' standardized test scores. This willingness has a limit, however. These researchers will most likely concede that a teacher is not highly effective if he or she consistently receives extremely low scores and does not seem to have any overly challenging external circumstances affecting the students in his or her class. In the same way, researchers who see teacher quality as determined by the results of a teacher's students standardized test scores tend to describe teachers as "highly effective" based on known characteristics which research has deemed to influence a teacher's effectiveness for students. These authors, as they assess teachers against standardized test scores and attempt to argue that effective teachers have significant influence on student achievement and growth, must rely on teacher qualifications as the basis of their distinguishing certain teachers at the outset of their research who they presume to be highly effective (Hanushek, et al. 1998). The research of these scholars tests which categories of teachers had larger impacts on student learning, measured by student performance on standardized tests. The basis of determining which teachers are "highly effective" and which are not, however, are the qualifications and characteristics teachers brought to the experiment before their students' achievement gains were measured in the study (Hanushek, et al. 1998). In this way, results-based researchers have also not wholly divorced themselves from the understanding that there are certain known teacher characteristics which influence effectiveness for students and which

increase and improve student learning. In fact, their acknowledgement of the importance of certain teacher qualifications, albeit not all the qualifications advanced as significant by qualifications-based researchers, is crucial to their argument.

Not surprisingly, these differences in approach to the question of the definition of teacher quality result in divergent policy preferences expressed by both groups. Qualifications-based researchers, tend to favor policies which recognize and reward teachers on the basis of pursuing higher-level content-area degrees or which ease restrictions on teachers who have served a given school district for a long time (Nye, et al. 2004, Clotfelter, et al. 2007, Harris and Sass 2007). Results-based researchers tend to emphasize the large impact teachers have on student standardized testing. Since, as they argue, effective teachers ought to see measurable, significant increases in student achievement with students in their classes, their research tends to be used in support of policies which evaluate teachers on their effectiveness in large part based on student standardized testing scores and tie these policies, although in a limited fashion, to school decisions about hiring, firing, tenure and compensation (Hanushek, et al. 1998). Researchers who see teacher quality as a results-based concept are more willing to describe teachers as highly effective who does not have the traditionally-valued qualifications, but who does, according to data, seem to have significant influence in bringing about gains for their students' achievement. These researchers tend to be more in favor, though support is certainly not ubiquitous in this literature, for programs which allow for alternative certification of persons not possessing the traditional high-level teaching qualifications. These researchers are more willing to accept the argument that teachers can be effective without having majored in education in college or without having secured graduate level degrees in education.

### The Use of Value-Added Modeling to Assess Teacher Quality.

As discussed, the most recent wave of education reform-related policy-making has framed the problem of poor student achievement within the U.S. public education system as the result of poor teacher performance and instruction in the classroom. This is why the proponents of tying significant portions of teachers' evaluation ratings to the performance of students on standardized tests, the results of which are sometimes used in decisions about teacher pay or tenure, argue that these measures are the most effective education reform policies. Policies that demand stricter accountability for teachers and data-based evaluation, they argue, can alleviate the issue of poor student performance because they target the very source of most of the problems with the public education system- poor teacher performance. Many of these claims are based on the research of scholars such as Eric Hanushek (Hanushek and Rivkin 2010), whose work has been used to support the claim that teachers are the most important factor in students' educations which schools can control and manage (Rice 2003). Early on, research identified a problem with holding teachers accountable for how individual students measured up against pre-determined "proficiency" scores on standardized tests (Darling-Hammond 2013). After all, there are many other factors which affect student academic performance which are completely outside teachers' realms of control, especially in school districts with high levels of poverty, high levels of racial or socioeconomic discrimination, low levels of early childhood development or nutrition, or low levels of parent educational attainment (Darling-Hammond, et al. 2012).

This reasoning gave birth to the creation of a new method of measuring teachers by their students' academic achievement, a method known as value-added modeling. Tennessee was one

of the first states to develop such a measurement system to assess student progress on its standardized tests, known as the Tennessee Value-Added Assessment System (TVAAS). As opposed to assessing students against absolute scores students were supposed to attain on tests, TVAAS, according to Tennessee's Department of Education, "follows the progress of individual students over time" (Tennessee Department of Education 2011). In this system, "each student is compared to his/her own past performance" and "each student serves as his or her own 'control'" (Tennessee Department of Education 2011). Thus, in states that utilize a value-added data assessment model, teacher effectiveness is defined by how much a teacher is able to improve upon student performance based on where the student was positioned academically at the start of the school year. Despite the admonitions of researchers like Haggai Kuppermintz (2003), value-added data is often used in districts across the country to make decisions related to the recruitment, hiring, firing, tenure and compensation of the nation's teachers. By last year, 38 states required that teachers be evaluated by measures of "the teachers' impact on student achievement" and 28 states required that school districts within the state "use teacher evaluations in making personnel decisions, such as tenure, promotions, license renewal, subject assignments or dismissal" (Lu 2013). Of these, a small, but gradually increasing number of states require the use of a value-added model in assessing student achievement data (Butrymowicz and Garland 2012). These several districts and states, however, are among some of the most influential in shaping education policy nationwide, including Chicago, Washington, D.C., Ohio and Florida (Butrymowicz and Garland 2012). Table 1 shows what some researchers have described as a "sea change" of teacher evaluation reform in recent years and it especially demonstrates the growth in the number of states adopting certain elements or formats of teacher evaluation since 2009,

following the start of the Department of Education's Race to the Top contest (Bornfreund 2013). In 2009, only fourteen states required annual observations of teachers and most of these only required annual observations for beginning or first-year teachers (Bornfreund 2013). By 2013, the number of states requiring annual observations had jumped to 28 and has continued to grow (Bornfreund 2013). The same observation is true of states requiring the use of student achievement data in teacher evaluations. Only fifteen states required use of this data prior to 2009 and they varied in terms of their definitions of "data" (Bornfreund 2013). As stated, as of 2013, over thirty-eight states now require that teacher evaluations include some measurement of student achievement data or growth as a factor in the yearly score teachers receive (Bornfreund 2013). What states require in terms of the significance this student data plays in teacher evaluations has also grown considerably. Prior to 2009, only three states required that student data be included as a "significant" component of a teacher's evaluation, yet by 2013, almost half of the states require that student data factor in as a significant part of teachers' assessments (Bornfreund 2013). Finally, the use of evaluations comprised of "multiple measures" or components, a concept which a number of researchers have endorsed and which was a recommendation of the Measures of Effective Teaching study, has also grown dramatically (Bornfreund 2013). Only twelve states had adopted evaluations comprised of multiple measures by 2009, but as of 2013, almost forty states have adopted such an evaluation format (Bornfreund 2013). This analysis of state teacher evaluation legislation is impressive because it represents the swift adoption of core components of teacher evaluation policy as envisioned by federal policymakers and private organizations such as the Bill and Melinda Gates Foundation within a relatively short period of time. This demonstrates the growing influence of national programs,



such as Race to the Top, in influencing state and local education policy and in shaping a relatively consistent and uniform “ideology of school reform,” of which teacher evaluations are a key item (Leana 2011).

## **Methodology and Approach**

Introduction. In attempting to offer a thorough analysis of the system just described, assessing its effectiveness in improving student achievement and in providing teachers with accurate, specific, helpful and constructive feedback to help them improve their performance and their contributions to student learning, this study takes three main approaches. It utilizes focus group conversations with teachers to help understand why teachers have the opinions they do about the system. It also uses limited data analysis to analyze the various scores given to teachers as a result of performance evaluations in the 2011-2012, 2012-2013 and 2013-2014 academic years. Because of the timing of this study, the 2013-2014 information only exists in the form of partial classroom observation scores given to teachers by principals. This information, paired with district demographic data allows for a thorough examination of the results of teacher performance evaluations in Memphis and Shelby County and, consequently, a better understanding of the effects they have and their implications for the system’s effectiveness.

### Teacher Focus Groups.

To gain even more insight into teacher perceptions about the fairness and effectiveness of the implementation and design of teacher evaluations, focus groups were conducted in November 2013 and April and May 2014. I conducted the first series of focus groups, which was sponsored by Shelby County Schools and which was jointly moderated and by staff of a third party research consulting group. All focus group sessions took place in a neutral environment provided by the third party research consulting group. All teachers were aware that Shelby County Schools staff would be moderating the focus group series and would be observing and recording the proceedings behind a one-way mirror. Teachers were also informed that Shelby County Schools staff would have full access to information about who participated in the focus group series and what was said during the meetings. Teachers were informed beforehand and gave consent to the recording of focus group sessions. Focus group times were offered on three evenings in mid-November after school hours.

Teachers who were asked to participate came from a random sample of 40 High School teachers, 40 Middle School teachers, and 40 Elementary School teachers who worked for the unified district. Some of the email addresses generated in the random selection process belonged to employees who had since stopped working for the district, resulting in a total sample size contacted of slightly less than 120 teachers. Teachers selected were randomized by which grade level they worked for- elementary, middle or high school. After receiving a mass email sent to all randomly selected participants, 20 teachers participated in the focus group discussions over the course of the three evenings. Teachers signed consent forms and answered short demographic

questionnaires before being provided with dinner and participating in the focus group's roundtable discussion led by either third party researchers or myself. The demographic questionnaire asked for participants to list the Shelby County Schools region<sup>9</sup> in which their school was located, their race, gender, grade level of their school (elementary, middle or high), the subject area in which they taught, and whether or not they would like to be contacted again for future research. The breakdown of teachers from each Shelby County Schools region was as follows: Northwest Region, 10% of participants; Northeast Region, 25%; Southwest Region, 10%; Southeast Region, 25%; East Region, 30%. Thirty-five percent of the participants in the focus groups taught in elementary schools in the district, while 40% taught in high schools and 25% taught in middle schools. Ninety percent of the teachers were female; 55% identified as African American, while 45% identified as white. Questions asked in these focus groups were developed by Shelby County Schools staff and pertained primarily to teacher awareness of tools the district was using at the time to better inform teachers about the various elements of the teacher effectiveness program. The questions were not entirely aimed at the material which I was attempting to examine in the course of this study, however, some of the feedback teachers provided during these working groups was valuable to the discussion of teacher perceptions of the fairness, implementation and design of the teacher evaluation system currently in place and being used by Shelby County Schools.

A second focus group series began in April 2014 and focus group times were offered for teachers to participate at Rhodes College during five evenings in late April. Later, an online form

---

<sup>9</sup> Prior to the school merger, Memphis City Schools operated based on a division of its schools and employees into four regions, Northwest (NW), Northeast (NE), Southwest (SW) and Southeast. Following the merger in July 2013, the unified district, Shelby County Schools, added a fifth region, the East (E) region, which comprised of the schools and employees added from the old Shelby County Schools system.

was opened for teachers to respond on their own time through May 15th. I designed all questions and moderated all group discussions. Teachers who opted to participate were offered entry into a drawing for two \$10.00 gift certificates to Starbucks. Teachers participating online and in the focus groups were asked to complete a brief 10 question, multiple choice demographic questionnaire which asked them to consent to participating in the study. Questions asked covered teacher school, SCS region, subject area, grade level, race, gender. Following their completion of the 10 question demographic questionnaire, teachers were asked to respond to 11 questions or statements. Teachers had the option to provide written answers or to reply via multiple choice options to save time. Thirty-nine elementary school, forty middle school and thirty-nine high school teachers were contacted after their names were generated in a 118 teacher randomized sample. Each grade level category of teachers was randomized by region of Shelby County Schools, so equal teachers from each region of the school district were contacted. To fully assess the working of the teacher evaluation system and specifically issues arising as a result of principal classroom observations and rating of teachers, I initially requested to conduct focus groups with district principals. The district provided a randomized list of 40 elementary, 40 middle and 40 high school principals and stated that focus groups for principals would be permitted once further approval was granted. Three principal focus groups were scheduled after school hours at the end of April, however the district ultimately decided not to grant the request and ultimately did not permit principals to be contacted and sign up for participation in focus group sessions. Of the 30 teachers who participated in the focus groups, roughly 40% taught in elementary schools, 30% taught in middle schools, and 30% taught in high schools in the district. Teacher participation by region was as follows: NE, 27%; NW, 20%; SE, 23%; SW, 20%; and E,

10%. Over eighty-six percent of the participants were female; 53% of the participants identified as white and 40% as African American.<sup>10</sup> The remainder of the participants preferred not to respond. Over seventy percent of the participants said that they taught at schools where the majority of students were economically disadvantaged. Slightly over sixteen percent of respondents did not teach students who were economically disadvantaged and ten percent did not know the number of students at their school who were economically disadvantaged.<sup>11</sup> Almost ninety percent of teachers who participated in the focus groups said that they taught in schools serving primarily students who were racial minorities. This was not the case for about ten percent of teachers who participated in the study. An additional question on the demographic questionnaire asked teachers to self-report their 2012-2013 summative evaluation score, if they were comfortable doing so. While slightly under 7% of respondents opted not to respond to this question and while the same percentage reported that they did not have a 2012-2013 summative score (most likely new teachers in the district), the number of teachers who did respond to this question was significant. Roughly 40% of the teachers who responded reported a 2012-2013 summative score of 4, 30% reported a summative score of 5 and almost 17% reported a 2012-2013 summative score of 3.

---

<sup>10</sup> Two participants (6.67%) in this sample preferred not to respond on the questionnaire in regard to the subject of race.

<sup>11</sup> For the record, whether or not a student is economically disadvantaged is determined based on whether or not the student qualifies for federally reimbursed free or reduced-price lunches, which is determined based on if the student's family falls within 130 to 185 percent of the federal poverty threshold.

### Data Analysis.

Much of the study involved analysis of data requested from and provided by the Shelby County Schools. Information requests were approved and granted by the district's Offices of Talent Management and Planning and Accountability. The district provided individual teacher evaluation data for all the district's 9,732 teachers who had received scores in the district's performance evaluation system for either the 2011-2012, 2012-2013 or 2013-2014 academic years. As stated earlier, the only information provided for the 2013-2014 academic year, due to the timing of the study, was the partial score received by the teacher for Fall 2013 for their classroom observations by their principals. Each teacher in the file was assigned a random identification number; teacher names were not included in the file. Information by teacher included the teacher's region, school name, partial observation rating for 2013-2014, summative score for 2012-2013, summative score for 2011-2012, TVAAS data score for 2012-2013, TVAAS data score for 2011-2012, observation rating for 2012-2013, observation rating for 2011-2012, student achievement data rating for 2012-2013, and student achievement data rating for 2011-2012. In addition, each year the Tennessee Department of Education publishes a report card which lists student achievement for the state, for local districts and even by school. This information is free and available to the public on the department's website. This information was accessed and used to track aggregate student achievement for Memphis City Schools and Shelby County Schools for the years since the teacher evaluation system was implemented (2010-2011, 2011-2012, and 2012-2013). Information provided on the department's website tracks district and school profile demographic information, student achievement information, student value-

added information, attendance and graduation rates, discipline information, teacher information, special education and Career and Technical Education (CTE).

### Research Questions.

The research questions for which I sought answers from my data are as follows:

- 1) Since adopting the teacher evaluation policy, have student achievement scores for Memphis City Schools, Shelby County Schools and the unified district risen, fallen or remained the same?
- 2) What are teacher perceptions of the teacher evaluation system? Do they report confidence in its accuracy, fairness, and effectiveness? Do they feel that it provides a solid standard for defining effective teaching? Do they feel that it offers them the tools necessary to improve on past performance? Do they feel that their evaluation scores are adequate and accurate reflections of their contributions to their students' learning? Do they understand all of its component parts and requirements? Do they understand the professional development and support opportunities that are open to them? Do they find these helpful?
- 3) Do teachers who have high scores from student value-added data (TVAAS scores) also have relatively high scores from the principal observation, student achievement, student perceptions and professionalism components of the evaluation summative score and vice versa (i.e., are the different components of the evaluation aligned with each other)? If not, what is the cause for any discrepancy?

- 4) What are the long-term effects of these teacher evaluation policies specifically on retention of teachers and on teacher morale? The answer to this question is especially important given the widespread implementation of teacher evaluation policy across the country and also the many high-stakes organizational decisions (such as those on hiring and firing, promotion, compensation, bonuses, etc.) in which they are used.

### Hypotheses.

In giving a preliminary answer to the above questions, I was sensitive to the fact that most of the qualitative information I relied on was provided by teachers, individuals who have a high stake in the outcome of this discussion since the topic at hand involves a discussion of their strengths or weaknesses as a teacher in an urban school district. Especially given the high stakes the state and district place on the scores resulting from performance evaluations, information from which can now be used in personnel decisions and in the issuing of organizational rewards and punishments, teachers who have received low scores are more likely to participate in the study and voice their opinion. Teachers who receive high or mediocre scores are less likely to participate and less likely to have a strong opinion. Still, I relied on the argument advanced by Carolyn Heinrich in her article entitled, “Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness,” published in the *Public Administration Review* in the November-December 2002 issue. In her article, Heinrich argues that whether or not employees generally perceive performance evaluation systems as fair or accurate measures of employee performance *can* have a significant effect on the success of



such systems in improving program output (2002). This argument squarely applies to Shelby County Schools' teacher evaluation program in place since the fall semester of 2010. Even though much of the data presented in this study reflects somewhat biased teacher opinions, I argue that this information is still worth considering especially given, as will be discussed later, its breadth, the intensity to which it is expressed and the fact that it is consistently expressed not only by teachers who receive low performance scores, but also by teachers who have received high scores who, in somewhat significant numbers, still fear that the performance evaluation system rates them on issues not within their control. The results presented below are complex and ought not be oversimplified. They provide a detailed and, at times, seemingly contradictory account of a system which is still very much a work in progress and which has many moving parts. To provide a performance evaluation system which both assesses in real time the performance of teachers in the classroom based on data both for the purpose of district decision-making and in order to provide teachers with effective, constant feedback and support designed to help them improve their instruction throughout the year brings with it a host of challenges and hurdles even for a small school district. When the 14th-largest school district in the country attempts to coordinate this formative and evaluative program for its more than 9,000 teachers and over 250 principals and schools, the attention to detail required of each individual who has a hand in the program's implementation is staggering and the execution, especially in the early years can be expected to be far from perfect. Keeping this in mind, I attempted to offer a fair and realistic preliminary answer to the questions posed in the form of my hypotheses. I hoped that this would result in an analysis of the effectiveness and implications of the policies implemented by the district to date.

Even with these limitations in mind, I still find Heinrich's argument to be a fair one, especially given that this element is particularly emphasized by the Gates Foundation in its approach to designing and implementing evaluation policies for the districts in which it works (The Bill and Melinda Gates Foundation 2014). The foundation stresses that teachers, principals, administrators, parents, students and the community ought to be included and involved in the ongoing work of teacher effectiveness, that frequent solicitations ought to be made to gauge their support for the work and that effort ought to be spent securing such support (The Bill and Melinda Gates Foundation 2014). Clearly, the Gates Foundation sees teacher perceptions of effectiveness and fairness of the system as important to the evaluation program's ultimate success. This is why I believe that my offering of a study which primarily relies on an understanding of teacher perceptions of and reactions to teacher evaluation policymaking in the district and the state still makes a worthwhile contribution to the literature on the subject. If teachers or principals are unsupportive or if, as shall be discussed further on, they feel as though the evaluation attempts to hold them accountable or punish them for factors which fall outside their control, they are likely to pursue two options. Teachers and principals, if they feel the system is unfair or ineffective, will either engage in activities aimed at skewing their performance evaluation score, thereby ensuring that the evaluation system does not provide accurate assessments of performance, or they will become significantly demoralized and attempt to leave the system altogether. In an urban public school district, one which has a history of poor performance and which has expended considerable resources to take significant steps toward improving teacher performance and retaining highly effective teachers, this is something that ought to be avoided. This is why a detailed consideration of teacher perceptions of system design

and execution must be continually kept up by the district. In its third year, with significant revisions to the rubric and to the way that teachers are trained on and utilize the evaluation process, such a detailed assessment of teacher perceptions of the system is warranted and worthwhile.

Specifically, in answer to the questions, given trends in the state and the district, I hypothesized that student achievement would increase in the district following the adoption of the policy. I do caution readers, however, against attributing this rise in student achievement scores solely to the teacher evaluation policy as Memphis, Shelby County and the State of Tennessee have all invested numerous resources in several different types of policies aimed at reforming education within their respective localities. This rise in scores, I argued, ought to be seen as the result of revived focus placed on education reform in the state, county and school district and not as the sole result of teacher evaluation policy changes, although this is most likely a factor. I hypothesized, in answer to the second question, that teachers have primarily either weak or negative perceptions of the evaluation system. For several reasons, I hypothesized that teachers would report relatively low confidence in the system's accuracy, fairness and effectiveness. Given teacher's considerable involvement in the research which led to the Teacher Effectiveness Measure rubrics, I posited that teachers, compared with these other factors, would have relatively positive perceptions of the rubrics and other conceptual frameworks as strong standards for defining effective teaching. I hypothesized that most of what teachers would offer in terms of negative feedback will center around the implementation and execution of the system. I argued that teachers would most likely understand the program's good intentions and its aims to provide them with effective performance feedback as well as to improve student

achievement. Nevertheless, I argued that, given the district's inevitable struggle to provide the resources and support it has identified that it needs to offer to teachers, the most significant of teachers' complaints would rest with program implementation and execution.

With regard to the data collected from the school district, I hypothesized that there would be strong correlations between the student value-added scores (TVAAS scores), student achievement scores, scores from principal classroom observations, student perceptions survey results and scores on the professionalism rubric. These elements have undergone significant research scrutiny and the skills which highly effective teachers are identified as having according to the TEM rubric are those same skills which, it is *assumed*, will bring about high student achievement results. I argued that the various multiple measures included in the teacher evaluation system used by Shelby County Schools are inter-reliable and correlate strongly. In other words, I hypothesized that teachers with strong value-added scores would also have relatively high scores from principal observations and student achievement data and vice versa. I hypothesized that the various multiple measures of the evaluation program would be aligned.

Finally, I hypothesized that there would be moderate negative effects of the system on teacher work and morale. This stems from the significant levels of negative teacher perceptions of the evaluation system, meaning that teachers have relatively low trust in whether or not the system is an effective assessment of their performance. When teachers feel as though they are being rated on factors which are outside their control, this can bring down morale significantly as has already been discussed in the literature review, and can result in teachers either leaving the district to work elsewhere or attempting to skew the portion of their performance results which they can control. Both of these options are not optimal for the district.

I hypothesized that teacher morale would be relatively low as a result of the implementation of the evaluation system and ongoing teacher concerns about the system which they feel have not been addressed.

### Approaching the Issues.

The process of teaching and learning, the process by which students learn, retain and apply information, the process by which students are assessed on how successfully they have mastered material and the shortcomings of our understanding of each of these processes make assessing teaching particularly complex. To assign weighted value to certain observations or measurable outcomes (i.e., student standardized test scores) of a teacher's performance and then to determine to what extent these factors ought to play a role in decisions about teacher compensation, bonuses, tenure, promotion, and retention, is certainly a very difficult business. To analyze this issue, to understand how student performance is improved and what factors influence the scores teachers receive on their performance evaluations is one which requires a holistic analysis that examines all the potential variables involved (Barton, et al. 2004). I hypothesize that there are many factors that affect the scores teachers receive on their evaluations. In addition, there are several factors which have the potential to translate this score into improved student achievement and teacher performance. In this way, I have identified *Teacher Classroom Performance, Principal Perceptions of Teacher Performance, Teacher Effect on Student Learning, Student Performance, and Student Perception of Teacher Performance* as the most significant independent variables. A host of antecedent variables, however, are also

involved and influence these independent variables. These include everything from teacher, principal, school and student demographics to a teacher's human capital (years of experience, drive for excellence, pedagogy and classroom skills, education level, content-area knowledge, reliance on and usage of available professional development and support opportunities, morale) or to principal's level of education and knowledge, understanding of the rubric or relationship with the teacher. Ultimately, because of the limitations of this study, I am not able to consider each of these variables and to conduct a full-scale measurement of the extent to which my hypotheses are correct regarding the placement of these variables into categories and groups which consequently have an effect on a teacher's evaluation score and, indirectly, on improving student achievement in the district, considering the system is used and implemented in a fashion which brings about these results. Instead, I will focus on assessing primarily *Principal Perceptions of Teacher Performance* in the form of the classroom observation scores teachers received in Shelby County Schools (the unified district, in 2013-2014) as well as in Shelby County Schools (pre-merger, 2011-2012 and 2012-2013) and in Memphis City Schools (pre-merger, 2011-2012 and 2012-2013). I will also examine teacher morale and teacher perceptions of the evaluation system which, I argue, are significant factors which inhere in the composition of both *Teacher Classroom Performance* and *Teacher Effect on Student Learning*.

It ought to be noted that there are a host of other issues which affect student learning in classrooms, particularly in a school district such as Shelby County Schools which serves a large and diverse population of students, some with extreme needs, who live below the poverty line and who come from households where parents or guardians have relatively low levels of formal education and limited resources. These contribute to what Barton et al. describe as the complex

network of variables involved in student learning and educational activity (2004). Of these other variables, it might be hypothesized that parental actions with regard to their students' educations form one of the most significant independent variables affecting student educational outcomes. It must be stressed, however, that a host of antecedent variables complicate the situation, however, especially involving the circumstantial factors already identified by other researchers, which include the age and grade level of the students, race (of both students and parents or guardians), gender (of both students and parents or guardians), the educational attainments of parents or other guardians who care for the student in question, household income level, and employment status of the parents or guardians involved. Location and local condition of the public school system is also an important factor. These variables, however, are prime examples of the difficulty one runs into when attempting to measure or assess their effect on a student's learning simply because they are so difficult to measure. The quality of a school district, for example, is a variable for which there is not any all-inclusive test or standard. The adequacy of schools and school systems in comparison with other schools and school systems is a contentious issue involving many different factors which has been thoroughly debated by scholars and policymakers alike. Even for a variable which might seem simple to measure or identify, such as family engagement, there are complexities that arise once the analysis starts. In measuring family engagement, for example, it is important to consider the types of family engagement sought after by parents and encouraged by the school district or school system which are of crucial importance, not just the simple fact that family members are engaged (Pomerantz, et al., 2007). As Pomerantz, et al., (2007) noted: "To date, the research conducted on parents' involvement in children's education has generally taken the approach of examining the extent to which parents

are involved, with more involvement on the part of parents being better for children. Although such an approach a fundamental first step, factors beyond the extent of parents' involvement are of major significance.” Other scholars have noted that, when it comes to family engagement specifically in relation to time spent completing homework, it is “more than minutes” that count (Epstein and Van Voorhis, 2001). For this reason, I choose to focus my study on the variables listed above and to attempt to derive conclusions from the data publicly available through the State of Tennessee’s Department of Education and that which has been specifically provided by Shelby County Schools’ Office of Planning and Accountability for use in this study.

## **Discussion of Results**

### Student Achievement in Memphis and Shelby County Since the Adoption of the Teacher Effectiveness Initiative in 2010.

Tables 2 and 3 demonstrate a brief look taken at student academic achievement, particularly that of 3rd through 8th graders on math and reading, in Shelby County Schools and Memphis City Schools from 2004 to 2013. In each chart, the line charted in between the darkest (Shelby County Schools) and lightest (Memphis City Schools) represents the state average for each year for which data was collected. This data is available on the state’s Department of Education website; it is provided in the form of the state’s annual report card which it releases each year (Tennessee Department of Education 2013). Each year the state administers the Tennessee Comprehensive Assessment Program or TCAP to students across the state. All students who are



enrolled in Tennessee public schools, grades 3 and older, sit for the examination, which is administered at the end of each school year. The state's use of this testing regimen was adopted and implemented as a result of the No Child Left Behind Act of 2001. Interestingly, the 2013-2014 administration of the TCAP marked the perennial use of the test to assess student achievement (Ujifusa 2014). Beginning in academic year 2015-2016, Tennessee will begin using a test created by the Partnership for Assessment of Readiness for College and Careers or PARCC test (Ujifusa 2014). This test is aligned with the new "Common Core State Standards" (Ujifusa 2014).

Table 2 shows growth in student academic achievement on the TCAP mathematics assessments for 3rd-8th graders from 2004-2013 for Memphis City Schools (pre-merger), Shelby County Schools (pre-merger) and the Tennessee average on the mathematics assessment. It is worth noting that, in 2009, Tennessee transitioned to "new curriculum standards and assessments more reflective of national and international student performance in the 21st century" (Tennessee Department of Education 2013). This is what accounts for the precipitous drop in student achievement for the state and both school districts between 2009 and 2010. I include years prior to the 2009 transition in assessments, however, because 2009 marks the beginning of Tennessee's attempt to seek recognition for its education reform initiatives through the Race to the Top program. It also marks the announcement of Memphis City Schools' new teacher evaluation initiative and its intensive partnership with the Bill and Melinda Gates Foundation for the purpose of designing new, rigorous teacher evaluations. Even with the dramatic drop in the percentage of 3rd-8th graders who were considered proficient or advanced in mathematics according to the TCAP assessments, a gradual trend can be ascertained in the student test scores

shown. For Memphis, Shelby County and the State of Tennessee, test scores were gradually on the rise both prior to and following the 2009 standards and assessments transition. While Shelby County Schools students consistently performed higher than those of Memphis City Schools students, it is worth noting that students from both districts were improving gradually.

While they do not exactly mirror the results shown in Table 2, the percentage of students who attained a ranking of “proficient or advanced” on the TCAP 3rd-8th grade Reading/Language Arts assessment between 2004 and 2013 for Memphis City Schools, Shelby County Schools and Tennessee follow a similar trend. On the whole, reading scores for all three entities are less consistent, but, overall, a general tendency toward growth and improvement is demonstrated both before and after the 2009 transition. It should be noted that after 2011, both Shelby County Schools and the state’s average score experienced a decline and then a gradual rise up through 2013. Memphis City Schools, however, experienced a decline in the percentage of its students who were ranked as proficient or advanced in reading and language arts from 2011 to 2012 and its average score did not experience significant growth following this decline as of 2013.

It must be stressed that the conclusions which may be drawn from this data are extremely limited. Especially with the significant change in standards and assessments instituted by the state in 2009, comparisons across years during this period are tentative at best. Even the Department of Education, on its report card, noted for these years, that “the 2009 achievement scores and all grades connected with these scores are considered the new baseline for future public reporting” (Tennessee Department of Education 2013). The department also noted that the “converted achievement scores” for 2010 onward were “based on restructured calculations and a redefined grade scale that are updated to reflect the current status of educational achievement in

the state” (Tennessee Department of Education 2013). Because of these dramatic transitions in how test scores were ranked, as reflected in the two tables, the state claimed that the “2009 change has prohibited comparisons to previous years’ data for achievement reporting including state, district, and school-level scores and grades” (Tennessee Department of Education 2013). Other limitations of the data are obvious. Because of the limited nature of this report, a complete analysis could not be conducted for all measurements taken of students achievement in the state. Notably, TCAP assessments administered to students enrolled in the 9th, 10th, 11th or 12th grades were not considered. Trends evident in graduation, drop-out, and attendance rates, discipline information, ACT scores and, perhaps most notably, Tennessee’s value-added student data, were not considered. All of these factors are considered vital measures of the health of districts and schools in Tennessee and yet a complete analysis of all this information could not be conducted. In addition, the choice of 2004 as the start date for analysis is relatively arbitrary and is due to the fact that Tennessee significantly altered its format for reporting student achievement data between 1997 and 2013, the years for which the state’s annual report card are available. With more time and more information, a more complete analysis of all these factors for longer periods of time ought to be conducted. Given the constraints of this report, however, such a thorough analysis was not conducted.

The question I attempted to answer in gathering this data was as follows: Since adopting the teacher evaluation policy, have student achievement scores for Memphis City Schools, Shelby County Schools and the unified district risen, fallen or remained the same? I hypothesized that student achievement would increase in both districts following the state’s win in the Race to the Top competition and following Memphis’ acceptance as an intensive partner site for the Bill and

Melinda Gates Foundation. I did stress, however, that attributing this rise in student achievement scores solely to the adoption of teacher evaluation policies in the district and in the state would be improper, due to the fact that Memphis, Shelby County and the State of Tennessee have all invested increased attention and resources on education reform since 2009. They have each also engaged in numerous policymaking efforts aimed at reforming education within their respective localities. Tennessee, for example, adopted a number of significant policy changes as a result of its Race to the Top proposal. While teacher evaluation policy was a significant component of that proposal, it was by no means the only change made to the way education was administered in the state. Tennessee also took steps to move toward the transition to the Common Core curriculum, created a state-run Achievement School District designed to take over and privately manage poorly performing schools in local districts, and made significant changes in the years following the start of its First to the Top program to teacher licensure and certification laws (Locker 2010). Because of this, while Tables 2 and 3 demonstrate that student achievement for 3rd-8th graders in reading and mathematics was gradually on the rise between 2004 and 2013, I argue that this cannot be solely attributed to teacher evaluation policy. I further argue that it is extremely difficult to ascertain with any degree of certainty the effect adoption of teacher evaluation policy had on the slightly improving test scores. As the new teacher evaluation systems are only in their third year of development and as the state and the two districts engaged in a number of extensive and simultaneous education reform initiatives, it is most likely that any effect these policies have on student achievement will only be determinable in the long-run and, even then, only as a rough estimate of the indirect influence these types of policies have.

As Professor Daniel Kiel of the University of Memphis' Cecil C. Humphreys School of Law argued, in a personal interview, it is very difficult to ever determine the direct effects of introducing "best practices" policies in education on student achievement scores, for a number of reasons (Kiel 2014). For one, as is evident in this situation, the introduction of new teacher evaluation policies in Shelby County and in Memphis was not conducted in a vacuum. They were introduced at the same time as other new policies were introduced and were most likely influenced in their own implementation and administration by the ongoing effects of other past, present and future policies. The myriad of activity, research and testing of new policies occurring at the time of Memphis' and Shelby County's adoption of new teacher evaluation policies and at the time of the state's grant of federal education money makes pinpointing any specific benefits to student achievement extremely difficult. In addition, as Mr. Kiel noted, the first year of the teacher evaluation system was a pilot period (Kiel 2014). Scores were not recorded and not all teachers were subject to the new system. This diminishes the already relatively short period these policies have been in place. Mr. Kiel argued that policies aimed at indirectly benefitting students, as these claim to do, may take years for results to show (Kiel 2014). These policies have not been in place long enough to assess their direct effect on student achievement. Finally, Mr. Kiel also noted that policies are only as effective as long as they are effectively implemented. Each year since the adoption of the Teacher Effectiveness Measure, the system has undergone significant revisions and reforms, most of which have been focused on its implementation. It is difficult to determine the effect of a policy which itself has undergone such significant changes in its own format and implementation since it was adopted. Even with all of these issues taken into account, there is significant reason to question whether or not teacher evaluation policies, while they may

still be desirable, will ever have any measurable effect on student achievement. Perhaps, as some researchers argue, the benefits are too intangible and too indirect to detect (Memphis-Shelby County Education Association 2014). While this is not necessarily an argument against the adoption of teacher evaluation policies at all, it does signify that an examination of student achievement since the adoption of teacher evaluation policies in Memphis and Shelby County is, at this time, premature and inconclusive.

Teacher Perceptions of the Teacher Evaluation System and Its Effectiveness in Memphis, Shelby County and the State of Tennessee.

Perhaps the most significant portion of the research collected as result of this report focused on teacher perceptions of system effectiveness, accuracy and fairness. As stated, this information came from teacher responses in focus groups with teachers conducted in November 2013 and April 2014.<sup>12</sup> The research questions which these responses were used to answer focused on teacher perceptions of the evaluation system itself, but also of its implementation. Questions focused on teacher perceptions of the system as a whole, of the various component elements included in their summative score, on their relationship with their principal, on the effectiveness and availability of professional development opportunities and on their willingness and frequency in taking advantage of these opportunities. The research questions addressed included those listed below: 1) What are teacher perceptions of the teacher evaluation system?, 2) Do they

---

<sup>12</sup> In order to ensure the authenticity of teacher responses and the freedom of teachers to respond to focus group topics, teachers in both focus group series were promised confidentiality. Participants were assured that no information provided in this report could ever be linked to individual teacher names, positions, schools or even regions. Because of this, no direct quotes are provided in this report, but only general concepts and ideas presented by the teachers.

report confidence in its accuracy, fairness, and effectiveness?, 3) Do they feel that it provides a solid standard for defining effective teaching?, 4) Do they feel that it offers them the tools necessary to improve on past performance?, 5) Do they feel that their evaluation scores are adequate and accurate reflections of their contributions to students' learning?, 6) Do they understand all of its component parts and requirements?, 7) Do they understand the professional development and support opportunities that are open to them?, and 8) Do they find these opportunities helpful? The results were somewhat consistent with what I had hypothesized. I had argued that teachers' opinions of the teacher evaluation system would be primarily negative, but that, for some components, teachers would have relatively mixed or high opinions of the system. I argued that teachers would report somewhat low confidence in the system's accuracy, fairness and effectiveness. This would contrast, I argued, with teachers' perceptions of the rubric and other conceptual frameworks, which would be relatively positive, given how involved teachers have been in the gradual construction and development of these systems and terms. I argued that teachers would focus on the Teacher Effectiveness Measure as accurately defining a standard for effective teaching and that teachers' negative perceptions of the system would be concentrated around the use of testing data, particularly that of TVAAS, as a portion of a teacher's summative score, and also around issues facing the implementation of the system. I had argued that, given the district's inevitable struggle to provide the resources and support it has identified that it needs to offer to teachers, the most significant of teachers' complaints would rest with program implementation and execution.

In November 2013, teachers focused primarily on answering questions designed by Shelby County Schools administrators to ascertain how effectively the district communicated with

teachers about the evaluation system. There was, however, one question that was asked at the focus group meetings which allowed teachers to respond and give feedback about their perceptions of the system's effectiveness and fairness. During focus groups, the moderator asked participants to respond to the following statement: "*The main purpose of the teacher and leader effectiveness work is to improve student achievement*" (Shelby County Schools Teacher Focus Group Series 2013).

It is interesting to note how, overwhelmingly, teachers responded in a very qualified manner to the statement. Unanimously, teacher participants in the district's focus group series affirmed that the mission of the district's teacher and leader effectiveness work was to improve student achievement. Several teachers, however, took the opportunity to specifically clarify that that was the *intent*, but not the actual *effect* of the evaluation program. These teachers were very clear in ensuring that, despite the wording of the statement and its attempt to ask teachers about the *purpose* of the teacher evaluation framework, they responded to whether or not they felt the teacher evaluation program was *effective* in accomplishing this purpose. When pressed by the focus group moderator on why they chose to make this distinction, these teachers stressed that they felt as though the teacher evaluation program used by the district had the potential to harm some teachers and some students' achievement due to its emphasis on rating teachers according to student test scores. Several of the teachers who held this opinion affirmed that they felt this was an unfair position. The moderator of the focus group pressed teachers further, asking them to respond to whether or not they felt as though the district's emphasis on measuring teachers according to student test scores was appropriate, given the research demonstrating that teachers are the most important variable affecting student achievement. One teacher in particular,



although others echoed this teacher's sentiment, argued that the research on this subject had been misconstrued. The teacher sought to emphasize an important distinction. Teachers, the teacher argued, are, according to the research, the most important variable *that the school can control*. This teacher, however, felt as though the qualifying condition- that the teacher is limited to the category of variables which the school can control and that there are a host of other variables which affect student achievement which are outside the school's and the teacher's control- had frequently been left off in discussions about a teacher's effect on student learning. This oversimplification of the research, as the teacher framed it, often served as a justification for evaluation programs like that adopted by Shelby County Schools which attempted to measure a teacher's effectiveness according to his or her students' performance on standardized tests. The reason, this teacher argued, that the teacher evaluation system in Shelby County did not fulfill its intended purpose, was because it unfairly held teachers accountable for performance of students on standardized tests, an outcome which the teacher certainly affects, but which is far beyond his or her control. Many of the teachers who participated in the focus group series agreed, at least to some degree, with this teacher's assessment of some of the major problems facing teacher evaluations in the district. While teachers echoed this sentiment throughout the focus group series, many of the questions were not applicable for use in this study and teachers' responses to this question effectively encapsulates their most general response to the teacher evaluation system adopted by Shelby County Schools. (Shelby County Schools Teacher Focus Group Series 2013).

In April 2014, I conducted a separate focus group series, during which I was the sole moderator of the focus group conversation. Focus group sessions were held at Rhodes College

and no Shelby County Schools staff were present, unlike the focus groups conducted in November. During these focus groups, as previously stated, teachers were asked to give a brief set of demographic information before proceeding on to answer questions about their perceptions of various elements of the teacher evaluation system in general. Teachers were asked to respond to a set of statements regarding the teacher evaluation system as a whole and the TEM rubric specifically. These statements attempted to ascertain teacher perceptions about whether or not the system and the rubric provided an effective, general standard for defining effective teaching in the district, one of the initial aims of the program. Teacher views regarding both the system and the rubric were relatively mixed, although the rubric tended to be viewed slightly more favorably by teachers, in accordance with my initial hypotheses. Teacher participants responded to the following statement regarding the system, *“The teacher evaluation system at Shelby County Schools as a whole (rubric, principal observations, use of student testing data, and student perception surveys) provides an accurate standard for defining effective teaching.”* Close to fifty percent of teachers either disagreed (30%) or strongly disagreed (just under 17%) with the statement. Interestingly, twenty percent of teachers took a neutral position. Almost 34% of teachers either agreed (just under 27%) or strongly agreed (just under 7%). It can be noted that slightly more teachers disagreed than agreed and 10% more teachers strongly disagreed than strongly agreed with the statement. On the whole, however, I emphasize that the teachers’ response to the statement was slightly more negative than positive, but generally mixed. I hypothesized that teachers would be generally supportive of the system and the results of the teacher focus group series were slightly more negative than I had predicted. (Shelby County Schools Teacher Focus Group Series 2014).

Regarding the rubric itself, known as the Teacher Effectiveness Measure, teachers were asked to respond to the statement, “*The Teacher Effectiveness Measure (TEM) rubrics provide an accurate standard for defining effective teaching.*” Almost forty-seven percent of teachers either disagreed (33.3%) or strongly disagreed (13.3%) with the statement, slightly more than I had predicted. Exactly the same number of teachers (33.3%) agreed as had disagreed, but the number who strongly agreed (just under 7%) with the statement was roughly half the size of the number who had strongly disagreed. The number of teacher participants who took a neutral position (13.3%) was lower than the number of neutral teachers who responded to the question about the teacher evaluation system in general. Altogether, teachers were more evenly divided between those who saw the rubric as an “accurate standard for defining effective teaching” and those who did not than they were divided for the question about the system more generally. This resulted in fewer neutral responses and a more even split between those who agreed or strongly agreed and those who disagreed or strongly disagreed. This was close to what I had hypothesized regarding the rubric, although again it represents a slightly more negative response than I had anticipated. This is surprising given that the question focuses solely on the rubric, an item which teachers have been involved in shaping and which is supposed to represent a comprehensive consolidation of best practices knowledge in terms of teachers. It has been suggested that teachers would be more strongly supportive of such a rubric because it was designed by and for practitioners in their field (Finch 2012). (Shelby County Schools Teacher Focus Group Series 2014).

The next set of questions asked teacher participants to focus on several of the individual components of the teacher evaluation system: 1) TVAAS data, 2) classroom observations by a

teacher's principal, and 3) the results of the teacher's student perceptions or Tripod survey. The focus of these questions was to ascertain teacher perceptions of their own individual scores in these areas, rather than asking them to reflect abstractly on their opinions of the fairness of these components in general, regardless of any direct effect on their summative score personally. Each statement asked teachers only to react to their score for the 2012-2013 school year. The first statement to which teachers were asked to respond was given as follows: "*My 2012-2013 TVAAS score accurately reflects my full contributions to my students' learning during the year.*" The results in answer to this question were more positive than I had predicted. Just under 37% percent of teachers either agreed (23.3%) or strongly agreed (13.3%) with the statement. In contrast, roughly 43% of teachers either disagreed (23.3%) or strongly disagreed (20%). Twenty percent of teachers took a neutral position, meaning that the results were very evenly divided between all five potential positions. Slightly more teachers approached the statement negatively than positively, which was different from what I had hypothesized. I assumed that teachers would strongly oppose the view that their TVAAS score reflected their "full contributions to" student learning. Part of this is possibly explained, as noted earlier, by the wording of the statement, which asked teachers to respond to whether or not their individual TVAAS score reflected their contributions as a teacher. According to the demographic questionnaire, the overwhelming majority of the teachers self-reported their 2012-2013 summative score as either a 4 or a 5. This might shed some light on why teachers seem to be evenly divided between affirming or negating their TVAAS score for 2012-2013 as an accurate reflection of their teaching. Perhaps if teachers were asked in either a future focus group session or a future survey about their opinions about whether or not TVAAS scores *in general* offer an accurate reflection

of a given teacher's contributions to student learning, they might respond more negatively in accordance with my hypothesis. Given that the use of TVAAS scores has been the most negatively viewed by teachers and critics of evaluation systems in the past, further research in this area might help to alleviate this distinction (Ravitch 2012, Memphis-Shelby County Education Association 2014). (Shelby County Schools Teacher Focus Group Series 2014).

Teachers were even more positive, on the whole, regarding whether or not the score given to them following their principal's observations of their classroom performance, "accurately reflected" their "performance as a teacher." Interestingly, close to 17% of teacher participants strongly agreed with this statement while the plurality of teachers, almost 37%, agreed. The number of teachers taking a neutral position was still relatively high, but lower than for other statements, at just over 13%. Altogether, 33.3% of teachers approached the statement with a negative perspective, either disagreeing (30%) or strongly disagreeing (3.3%) that their classroom observation score accurately reflected their performance as a teacher. This was the first question in the focus group series to which teachers responded more positively than they did negatively. The explanation for this, which, again, was more positive than I had initially predicted, might lie also with the wording of the statement. As the data will illustrate, teachers are generally rated well by principals. Very few teachers are rated as 1s or 2s on their classroom observation portion and the vast majority are rated as either 4s or 5s (See Tables 6 and 7). In 2013-2014, for example, according to data provided by Shelby County Schools, over eighty percent of teachers were rated as either 4 or 5.<sup>13</sup> Perhaps if teachers were asked to reflect more

---

<sup>13</sup> It ought to be noted that, as data from the 2013-2014 school year will not be available until the 2014-2015 school year, this information is only taken from a partial report of classroom observation scores provided by Shelby County Schools in April 2014.

abstractly on whether principal ratings of teacher classroom performance in general serves as an accurate reflection of actual teacher performance, the results would be more negative. It is equally possible, however, that they might not, given teachers' stronger support for the rubric and for the observations portion of the evaluation program than for the other portions. Performance against the rubric, even though it is being scored by the principal, is something that teachers often see as within their control and as something which they can tangibly work to improve (Memphis-Shelby County Education Association 2014). Because of this, it is common for teachers to memorize and practice teaching according to the expectations of the Teacher Effectiveness Framework (Memphis-Shelby County Education Association 2014). Perhaps this, even more than the wording of the statement, accounts for why teachers tend to be more supportive of viewing their principal's scoring of their classroom performance as an accurate reflection of their performance. If teacher perceptions of problems with the classroom observation portion of the evaluation program, if any exist, are to be better detected, a more specific set of questions and discussions will need to occur. These will need to attempt to draw out from teachers specific factors which might lead to their *not* viewing principal scores as effective indicators of their actual performance. Future research will need to focus on determining if items such as teacher years of experience and education (which affect the number of times teachers must be observed by their principals in a year), strength of teachers' relationships with their principals, strength of teacher working environment and relationships with other teachers, and strength of teachers' understanding of the rubric, the concepts it attempts to measure and a knowledge of how to demonstrate effective knowledge and practice of effective teaching according to the rubric have an effect on how positively teachers see their classroom

observation scores as accurate reflections of their teaching performance. (Shelby County Schools Teacher Focus Group Series 2014).

Teachers responded much more negatively to the issue of student perceptions surveys, as I hypothesized. This has been, along with the use of TVAAS data, perhaps one of the most controversial portions of the teacher evaluation system and one of the most negatively viewed among teachers. It is not surprise, then, that teachers would respond more negatively than positively to the statement: *“The results of my students’ 2012-2013 perceptions survey are an accurate reflection of my performance as a teacher.”* Over forty-three percent of teachers either disagreed (33.3%) or strongly disagreed (10%) with the statement. The plurality of teachers disagreed and many more teachers had a strong negative reaction to this statement than to many of the other statements discussed during the focus group sessions. Interestingly, many more teachers also strongly agreed with this statement (13.3%) than with many of the other statements. Just under 17% of teachers agreed with the statement. This statement also drew one of the largest neutral responses in the course of the discussion. Almost a third of the teachers participating in the discussions took a neutral position as to whether student perceptions surveys served as an accurate reflection of their performance. Perhaps this is the result of new teachers in the system who have not yet ever had students who were given a perceptions survey and thus had nothing to report on. Perhaps it is the result of a considerable portion of teachers who have not formed any strong opinions about the survey due to the fact that student perceptions only account for 5% of a teacher’s summative score. (Shelby County Schools Teacher Focus Group Series 2014).

The next two questions which the focus groups discussed attempted to measure teacher perceptions, albeit in a limited fashion, of the day-to-day implementation and maintenance of the

teacher evaluation system. As established in the literature review, a performance evaluation system's ability to encourage and facilitate an organic, ongoing system of informal performance feedback and communication is, according to some researchers, one of the most significant benefits of performance evaluations. If, a performance evaluation system fails to achieve such a system of feedback, many researchers argue that it has failed to fulfill its highest purpose and is therefore not worth the costs of design and implementation, as discussed (Marlinga 2006, Nickols 2000). Thus, teacher participants were asked to respond to the following statement:

*“Through performance evaluations, my principal and district administrators provide helpful and constructive feedback that I use to improve my teaching.”* Teacher responses to this statement were intended to convey teacher perceptions about the effectiveness of the system at providing them with real-time, effective feedback on how to improve their practice, a stated goal of the teacher effectiveness system. This statement drew strong positive responses from teachers, which goes against my earlier hypothesis that the majority of teachers’ negative perceptions of the system would center around implementation and communication. Sixty percent of teachers reacted to this statement positively, with fifty percent agreeing and ten percent strongly agreeing that the principal and the district used the system effectively to provide helpful feedback. The statement was even worded to attempt to capture multiple teacher viewpoints with regard to feedback. Teachers were not only asked to respond to whether or not principals and the district provided effective feedback, but also whether or not this feedback was such that teachers could *actually use* it to improve their performance. No teacher strongly disagreed with the statement and only twenty percent of teachers disagreed. Interestingly, twenty percent of teachers responded neutrally to the question. Again, it would be interesting to delve more deeply into this



topic in particular, testing teacher responses for different items which may affect the ability of a principal or district administrator to give effective feedback through the performance evaluation system and the ability of a teacher to receive and implement such feedback to improve his or her performance. Future research may also attempt to measure teacher's views about the use of principals as providers of performance feedback (in the course of the performance evaluation cycle) more generally as opposed to asking to teachers to reflect on their individual principals or supervisors in their own workplace environments, to which they might have strong personal connections, even if principals or administrators provide weak performance feedback. Finally, there is an important element of communication and feedback which this statement, as it is worded, ignores. As mentioned earlier, performance evaluations are valuable insofar as they serve to generate authentic and accurate performance feedback. Several researchers, however, stress that this cannot be limited to the mandatory performance evaluation-related conversations required by the system itself (Marlinga 2006, Nickols 2000). Instead, these researchers argue, performance evaluations, if they are to be successful at actually improving performance, must encourage an organic culture of feedback and communication which exists independently of the formal performance evaluation system. Performance evaluations, some researchers argue, ought merely to be a tool to spark a culture of feedback which is consistent, ongoing and continuously present. This statement, and the responses it generated from teachers in this focus group series, however, did not measure the extent to which teachers felt that district administrators or their principals were giving them feedback on a regular basis, outside the confines of the performance evaluation system. The statement specifically asks for teacher's reactions to "helpful and constructive feedback" provided "through performance evaluations." This is an area into which

future research ought to investigate further. (Shelby County Schools Teacher Focus Group Series 2014).

The second statement in the portion of the discussion focused on communication and implementation asked teachers to reflect on the professional development opportunities offered by the district to assist teachers and help them to improve. The district has long stressed that, in describing its teacher evaluation system, it is intended to truly improve teacher performance, not merely to identify and punish those teachers who perform poorly (Shelby County Schools 2013). This is especially crucial for first-year teachers who, often, are still attempting to learn *how to teach* while also being responsible for the learning and achievement of an entire classroom of students (Shelby County Schools 2013). In order to gauge teacher perceptions of the effectiveness of the highly detailed and extensive network of professional development opportunities to which struggling or low-rated teachers are referred, teacher participants were asked to respond to the following statement: “*I understand the professional development (PD) and support opportunities that the district offers AND have taken advantage of these opportunities to improve my performance.*” The statement attempted to gain insight into teacher perceptions of how well teachers understood and have been able to take advantage of professional development opportunities. It also attempted to, but in a more limited fashion, gauge the effectiveness of the professional development opportunities offered by the district by stressing that the opportunities which teachers discussed had actually fostered improvements in the teacher’s performance. Teachers, in responding to this statement in the focus group discussions were, on the whole, strongly supportive of district professional development opportunities. Almost eighty percent of teachers agreed (56.7%) or strongly agreed (20%) that

they understood, took advantage of and were able to improve their performance as a result of district professional development opportunities. Just under 17% of teachers expressed a neutral position toward district PD opportunities and just under 7% disagreed with the statement. No teachers strongly disagreed. The statement is weakest in its ability to measure the effectiveness of teacher professional development opportunities for individual teachers and in general. Teacher responses to this statement most appropriately constitute an assessment of how well teachers understand and can use district professional development opportunities. Future research ought to attempt to ascertain more thoroughly how effective teachers perceive professional development and district support aimed at helping them improve to be. Whether or not teachers take advantage of professional development tasks which are recommended to them is cataloged as part of their records (Shelby County Schools 2013). Teachers are not punished for not taking advantage of professional development opportunities, but when a teacher decides to grieve the teacher evaluation process as not being carried out properly, the adjudication pays attention to whether or not teachers took advantage of all the help that was offered to them during the performance evaluation cycle (Shelby County Schools 2013). This system has the potential to cause teachers to be very attentive to available professional development opportunities and to see it as very important that they understand and take advantage of all support offered to them. It places less emphasis, however, on securing authentic understandings of whether or not teachers see such opportunities as effective and helpful in improving their practice. Teachers might tend to see professional development opportunities as items to acquire for their resumes or as tasks to accomplish on a checklist instead of as lessons, programs and support which have a measurable effect on their ability to improve. Future research should attempt to place before teachers this

type of question or statement and attempt to gauge their reaction in a survey or focus group discussion. As it is, the results from this focus group series in terms of assessing the effectiveness of teacher professional development opportunities are limited. (Shelby County Schools Teacher Focus Group Series 2014).

The last major section of the focus group discussion asked teachers to reflect on the fairness of the performance evaluation system in general, and on the fairness of the use of test scores in their performance evaluations specifically. This question was inspired by the discussions engaged in by teachers and district staff during the focus group series for teachers conducted in November 2013. Teachers were asked to give their responses to the following question: *“Are there factors outside of your control that you think significantly affect your students’ performance and your overall performance score?”* Precisely ninety percent of the teachers responded “yes,” and ten percent with “no.” The question was specifically worded to gauge teacher perceptions of two main issues. First, teachers were asked to give their perceptions of whether or not there were factors outside of their control which affected their students’ performance rates. Second, and with more direct relevance to the topic at hand, teachers were asked to assess whether these factors which might have an influence on student performance had a “significant” effect on their “overall performance evaluation scores.” The answer from the teachers was overwhelming. The vast majority of teachers see factors which are outside their control as significantly impacting the results of their evaluation. This, as Heinrich argued, and as has been argued earlier in this report, can have considerable damaging effects for teacher faith and trust in the evaluation system to provide accurate assessments of their performance. Consequently, low teacher faith and trust in the system can have an impact on the effectiveness of the system itself and the information it

provides to district administrators about districtwide teacher performance. Teachers, principals and district administrators have the potential to be affected by inaccurate or skewed performance which occurs once faith in the system begins to wane. For example, if teachers feel as though they are being rated on items which fall significantly out of their control, they might attempt to skew other elements of their performance evaluations, such as the classroom observations, they might increasingly attempt to modify student answers to standardized tests in extreme cases, they might choose to grieve their performance evaluation processes more often, they might experience significantly decreased motivation or work ethic or, again in extreme cases, they might leave teaching for the district to the teaching occupation in general in increased numbers (Heinrich 2002). I argue that this portion of the teacher feedback component of this study is one to which Shelby County Schools and other districts attempting to implement similar performance evaluations ought to pay considerable attention. Even though these responses represent the thoughts and opinions of an arguably biased group of individuals toward the effectiveness of the performance evaluation system, they represent nonetheless, the experienced opinions of a group of practitioners with expertise in their field. They also represent the opinions of those who are subject to the confines and constraints of the Shelby County Schools performance evaluation system. If the vast majority of teachers, as the responses from this representative sample suggest might be the case, feel as though there are significant factors outside their control which influence their evaluation score, and if, as has been suggested, this represents a significant decline in faith and trust on the part of teachers in the evaluation system, this has significant ramifications for the effectiveness of Shelby County Schools' teacher evaluation system. Teachers were asked to provide examples of these factors which they see as significantly

affecting their evaluation score and their students' performance rates and the examples are those which have been discussed earlier in the report. Many teachers suggested that issues related to either parents or the "home environment" played a significant role in student performance. Other teachers suggested that personal, student-defined attitudinal issues played a role, with some teachers arguing that defiant students or students with a lack of respect toward authority could not necessarily be controlled or made to perform well by teachers. Aside from student behavior and home life, an area where both of these play a role- student attendance- was seen as one of the most significant by teacher participants. Many teachers remarked that they had observed students who were simply unwilling to complete, or even to begin, standardized tests, significantly dropping a teacher's overall score. School-level factors were also suggested as some teachers suggested that class size, over-testing, schedule adjustments, passing students who did not meet grade standards the year before, and school experiences in other grades or at other schools could have significant factors in determining student achievement and, ultimately, teacher's evaluation scores. Future research ought to attempt to examine more deeply the nature of these factors which teachers argue affect their scores and to attempt, if possible, to assess the varying levels of importance which teachers assign to each of the factors. Future research might also examine to what extent teachers see school-level factors, as opposed to out-of-school factors, as involved in shaping student achievement. As Barton et al. have established, however, it is unlikely that definitive percentages or numbers will ever be determined for each of the factors or inputs which influence student achievement (2004). The process of teaching and learning is intensely complex and there are almost innumerable variables and categories of variables which may affect each student's ability to perform. Indeed, it may very well be that for each student or for groups of

students, some factors weigh more heavily than others or even that, for some students or groups of students, some factors are extremely significant whereas they are a nonissue for others. Still, it would be worthwhile, if anything to assess how appropriate current weightings for teacher evaluations are, to continue to investigate this matter both from the perspective of those who practice education, the teachers, and in the form of other data analysis and research. (Shelby County Schools Teacher Focus Group Series 2014).

#### Reliability of the “Multiple Measures” of Teacher Evaluations.

As discussed, data analysis was conducted on teacher evaluation scores provided by Shelby County Schools’ Office of Performance Management from 2011-2012 and from 2012-2013. For these years, individual teacher scores were provided for the student achievement, value-added and observation components of the evaluation. Additionally, partial observation scores for 2013-2014 were provided. These scores were for all teachers currently working in the unified district. Information from 2011-2012 and 2012-2013 only included scores for teachers working for Memphis City Schools as district administrators could not locate the equivalent data for those years for the old Shelby County Schools system. Tables 4, 5 and 6, shown below, depict the distribution of scores (1, 2, 3, 4, or 5) between all teachers recorded in the data file for 2011-2012 (Table 5), 2012-2013 (Table 4) and the set of partial observation scores for all teachers in 2013-2014 (Table 6). These tables are provided merely to show how many teachers received each of the five scores on their summative evaluations for these years. The results are interesting, showing that the vast majority of teachers each year receive 4s and 5s. Teachers

receiving a score of 3 fall behind this group and the numbers of teachers receiving 1s and 2s are minimal, if existent at all.

Table 4 depicts all summative scores for teachers in Memphis City Schools in 2012-2013. In this year, 70% of teachers received either a 4 (31%) or a 5 (39%). It is worth noting that a plurality of the teachers recorded, received a score of 5. Twenty percent of teachers received a score of 3, while ten percent of teachers received scores of either 1 (3%) and 2 (7%). The results from the year before, 2011-2012, are even more heavily weighted toward 4s and 5s and away from 1s and 2s. In 2011-2012, as depicted in Table 5, a plurality of teachers, just under 42%, received a summative score of 5 for the year. Falling closely behind this group of teachers, almost thirty-five percent of teachers receiving summative scores this year received a 4 as their score for the year. Slightly more than the year before, just under 24% of teachers received a summative score of 3 for the year. The number of teachers who received a score of 1 for the year, however, was so small that it could not be depicted on the chart and only 0.10% of teachers received a summative score of 2. When these results are analyzed, it is worth noting that the self-reported scores of teachers participating in the April 2014 focus group series are relatively representative of the official data collected and reported for the 2012-2013 school year.

The only data available at the time of the final preparation of this report for 2013-2014 were partial observation scores, meaning a record of the scores given by principals to teachers on the classroom observation portion of their evaluation to date. The score for each teacher, however, is incomplete, as all teachers at the time this list was generated, had at least one additional classroom observation session that was incomplete and needed to be conducted by their principals. These partial observation scores for 2013-2014, however, reflected in Table 6, align



with the summative scores from other years and reflect yet another trend evident in the data from 2011-2012 and 2012-2013 provided by Shelby County Schools' Office of Performance Management- teacher scores on the classroom observation portion of the evaluations, rated by principals, tend to be even more heavily weighted toward scores of 4 or 5 than the other components in the evaluation, with the exception (as shown in Table 7) in 2011-2012, of student achievement scores. Close to half of the partial scores given by principals to teachers as a result of classroom observation sessions in 2013-2014 were 4s; roughly 36% of scores given were 5s. Only 18% of teachers received scores of 3 from their principals as a result of classroom observations to date at the time the data file was generated. Even fewer than in 2011-2012, only 0.02% of teachers received a rating of 2 and the number who received scores of 1 was even smaller and could not be depicted on the chart. According to this information, it is worth noting that over eighty percent of all partial observation scores given for 2013-2014 were either 4s or 5s. This is compared with roughly 77% in 2011-2012 and roughly 85% in 2012-2013. Generally, as the data reflects, it ought to be noted that most teachers receive 4s and 5s on all components of the evaluation. In addition, even more teachers, on the whole, receive 4s and 5s on the observation component of the evaluation than on other components of the evaluation.

While such results may seem surprising, they are actually in line with what other districts across the country have struggled with when implementing teacher evaluations using data derived from value-added modeling. When a teacher evaluation regimen based in large part on value-added student data was implemented in Florida in 2011-2012, for example, large percentages of teachers received scores of "highly effective" or "effective," equivalent to Memphis City School's 4 or 5 ratings (Yi 2013). In fact, in 2012-2013, roughly 98 percent of

teachers received scores of “highly effective” or “effective” and these scores were relatively consistent with scores from the year before (Yi 2013). It was not uncommon for districts to have no teachers who were rated as less than “effective,” as occurred in Palm Beach County (Yi 2013). The same results were seen in New York in 2013, when the state instituted its new evaluation program. Roughly 92% of New York teachers were rated as either highly effective or effective when its scores were released in 2013 (Bakeman 2013). New York’s teacher evaluations were based in large part on student performance on a new exam aligned to the Common Core curriculum (Bakeman 2013). In Chicago, a news report following the announcement of new evaluation scores noted that, “A first and partial glimpse of Chicago’s new evaluation system for public school teachers indicates fewer of them are considered ‘excellent,’ but the shift in numbers is not as dramatic as once thought” (Vevea 2013). Only one percent of teachers in Chicago were deemed ‘unsatisfactory’ as a result of the ratings, with the large majority of teachers rated as “excellent or superior” (Vevea 2013). Other states and localities implementing these types of evaluation frameworks have seen similar results (Vevea 2013).

In fact, of these school districts, Memphis is actually considered one with a much more even distribution than is found in other districts attempting to implement new teacher evaluations (Vevea 2013). Interestingly, despite the vast number of teachers who receive these ratings, teacher opposition to the use of either value-added modeling or the inclusion of student testing data as a very significant factor in yearly evaluation scores has not waned (Yi 2013, Bakeman 2013, Vevea 2013, Rich 2012). In fact, if anything, teacher opposition to new evaluation programs has only grown. Teachers seem primarily to oppose, as they have in Memphis, the argument that what they contribute to student learning can be reduced to student performance on

tests, especially when so many other parent-related, home-related and school-related factors are at play (Rich 2012). It seems as though even teachers who receive relatively high scores on their evaluations oppose the new evaluation frameworks for two primary reasons. First, teachers oppose the new evaluation frameworks simply because they are not based on accurate measurements of teacher performance, in the eyes of many teachers. Even teachers who have high scores might still be inclined to oppose the evaluations because they are fearful that the rating of their effectiveness could rise and fall with student testing which, as established by the previous section, teachers overwhelmingly see as outside their control for the most part. Secondly, teachers most likely tend to oppose the evaluation frameworks because they see the focus of the evaluation systems as mistakenly centering around identifying, punishing and eliminating poor performing teachers. When the results often come back showing that the vast majority of teachers are effective in their classrooms, teachers argue that the resources used on evaluations could have been better spent on support or professional development to improve their already strong performance and allow them to keep learning about how to be a stronger instructor. In New York, for example, the president of the local teachers' union was quoted as saying, "We've known for a long time that better than 90 percent of New York's teachers are effective or highly effective, and these numbers, based on really questionable data, don't give me any sense of reassurance" (Bakeman 2013). A member of the state's board of regents added, "What's going through my mind is: If we had put that toward [professional development], if we had put that toward supports, and not the 'gotcha'...approach, would our children be better off?" (Bakeman 2013). These are the same arguments formulated by the majority of teachers, as

demonstrated in the previous section devoted to teacher feedback in the focus group portion of this study.

Table 7 depicts what is commonly referred to by school district administration and state officials as “rater drift” (Shelby County Schools 2013). Rater drift essentially refers to the disparity between the numbers of teachers who receive 4s and 5s on their evaluations and the numbers of teachers who receive 4s and 5s on the TVAAS and student achievement portions of their evaluations. As has already been noted, a larger percentage of teachers are rated by principals as 4s or 5s than receive 4s or 5s for their summative scores or on the other components that factor into their summative scores. In fact, between 2011-2012 and 2012-2013, the gap widened to the point that state officials became involved, meeting with local district administrators to encourage them to take steps to reduce the number of teachers who received 4s and 5s on their evaluations (Shelby County Schools 2013). In the summer of 2013, district principals were instructed that the inflation of 4 and 5 classroom observation scores was due to principals not being able to detect when a teacher was putting on, in the local school district vernacular, a “traditional dog and pony show” (where teachers attempt to skew principal classroom observation ratings by performing during observations very differently from how they would perform during a normal day in the classroom) (Shelby County Schools 2013). District principals were also told that the large number of teachers rated as 4s and 5s for classroom observations did not align with the number of teachers who received 4s and 5s on their TVAAS scores and that this might also mean that some principals were struggling to give teachers low ratings because they were attempting to avoid having difficult, confrontational meetings with teachers about their performance (Shelby County Schools 2013).

A look at Table 7 demonstrates what the state and district are concerned about. In 2011-2012, roughly 77% of teachers received classroom observation scores of 4 or 5, while only 53% of teachers received equivalent TVAAS scores. The disparity for 2012-2013 was even greater, with 85% of teachers receiving scores of 4 or 5 for classroom observations and only 59% receiving equivalent TVAAS scores. While depicted in the Table, the student achievement portion, which makes up 15% of a teacher's total evaluation, had a different relationship to classroom observations than did the TVAAS scores. In 2011-2012, 91% of teachers received 4s or 5s on their student achievement portion of the teacher evaluation compared with 77% of teachers on the classroom observations portion. In 2012-2013, the positions were reversed with 75% of teachers receiving 4s or 5s on student achievement, compared with 85% of teachers on the classroom observations portion. It must be noted that the disparity between the classroom observation and student achievement portions of the evaluation is not as great as that between TVAAS scores and classroom observation scores and that the relationship was actually reversed between 2011-2012 and 2012-2013. The student achievement portion of the evaluation is meant to be another measure of student learning, other than that provided by the value-added data, selected by the teacher and principal. The change from 91% to 75% is probably due to a policy change that was made between the two years. In 2011-2012, teachers were permitted to select from a district-approved list, which data would serve as the student achievement portion of their evaluation. In 2012-2013, due to complaints that teachers were selecting data most favorable to them, the district changed the policy to prescribe that teachers and principals would discuss and select the achievement data at the beginning of the year in a conference. If the teacher and principal disagreed, however, the principal had the final say in selecting the source of the

teacher's student achievement data. It is possible that this policy change had an impact on the shift in the number of teachers who were rated as 4s and 5s from 2011-2012, where this number was larger than the equivalent number for classroom observations, and 2012-2013, where this number was smaller than the equivalent number for classroom observations.

It is worth noting that the issue of rater drift is highly controversial with teachers across the district as is the state and district's approach to attempting to fix the discrepancy. From the perspective of teachers and the teachers' union, the state is making a largely unfounded assumption in asserting that the 53% (2011-2012) or 59% (2012-2013) of teachers who received a TVAAS 4 or 5 is the *actual* number of 4 and 5 teachers as opposed to the 77% (2011-2012) or 85% (2012-2013) who received a classroom observation 4 or 5. Teachers argue that it is inconsistent for the district to put faith in classroom observations when structuring the evaluation framework, weighting them anywhere from 40-55% of a teacher's evaluation, depending on the situation, in comparison with the 25-35% weighting given to TVAAS data, but then to choose the distribution of TVAAS 4 and 5 scores as the "correct" scores and asserting a problem with classroom observation scores. As already noted, many teachers do not see value-added modeling data, or even general student testing data, for that matter, as reliable indicators of actual teacher performance. They argue that the ratings given by principals, who have been extensively trained on how to rate according to the rubric, in the course of classroom observations may actually be the more correct indicator of teacher performance. Representatives of the teachers' union in Shelby County, the Memphis-Shelby County Education Association, argued that because of the complexity of the teaching and learning process and because there is no clear, research-based determination that has established that a "highly effective" teacher, as measured against the

rubric, will have high value-added scores, the district ought to accept the principal observation scores as the more “correct,” if such a term can be ascribed due to the uncertainty of both measurements, in the face of a discrepancy (Memphis-Shelby County Education Association 2014). This is only fair, the union representatives argue, since the district has established the rubric and the ratings of principals, who have been extensively trained in rating teacher performance against the rubric, as the generally-accepted definition of effective teaching (Memphis-Shelby County Education Association 2014). If the district has opted to affirm the accuracy of TVAAS scores when they contradict percentages of scores, then why, union representatives argue, invest the extensive time and resources that have already been invested in developing, implementing and perfecting the rubric and the complex process of principal observations (Memphis-Shelby County Education Association 2014). For this reason, teachers and their union representatives assert that, in the dilemma caused by “rater drift,” the district should affirm the scores given by principals against the rubric as more accurate than TVAAS scores.

Even aside from teacher perceptions of the fairness or accuracy of the district’s trusting TVAAS scores as opposed to classroom observation scores given by principals in observing teacher performance against the TEM rubric, the issue of “rater drift” is one which poses problems for the integrity of the teacher evaluation framework used by the district. If there is so great a discrepancy between the number of teachers who are rated as 4s and 5s by principals and the number of teachers who receive equivalent scores according to TVAAS data, this means that there are problems that exist in the inter-rater reliability of the various items of the teacher evaluation system. In theory, a teacher who is considered highly effective because he or she

receives a yearly summative rating of 5 ought to be affirmed as 5, or at least as rating close to a 5, by all items which contribute to the summative rating. It is a signal of the weakness of the rating system used if the 5 is derived from the average of widely divergent scores provided by the various elements included in the evaluation or if the teacher only receives a 5 because he or she received a 5 in the most highly weighted element of the evaluation while receiving much lower scores on the other components. It is generally established that instances of rater drift are a sign of system error or of a need for more training by raters. To test the strength of the correlations between various elements of the evaluation system, I received assistance in running a correlation between the three elements of the evaluation system (classroom observation score, TVAAS score, and student achievement score) for each year for which data was provided by the district. I used Pearson's correlation coefficient  $\rho$  to measure pairwise correlation among the three elements of the evaluation as a means of assessing the inter-rater reliability of these elements. I used the "crude estimates" for interpreting the strengths of correlations between the elements provided by the Political Science Department at Quinnipiac University in determining the statistical significance of the relationships I uncovered between the elements.

What I found aligned with the discussion provided above regarding rater drift- that there is little to no relationship between each of the three primary elements used in evaluating teachers and that this can result in a significant discrepancy between various components of a teacher's evaluation. For 2012-2013, the value returned for the relationship between teachers' TVAAS data scores and teachers' classroom observation scores was 0.11, meaning that there is a high probability that any covariation occurred by chance and that there is either no statistically significant relationship between the two items or only a negligible relationship. For 2012-2013,



the value returned for the relationship between teachers' student achievement data scores and teachers' classroom observation scores was 0.10, also signifying either a non-existent or very weak relationship. For 2012-2013, the value returned for the relationship between value-added scores and student achievement scores was 0.52, meaning that there was a relatively strong positive relationship between the teachers' scores derived from student value-added data and scores derived by other measures of student achievement. This was the only statistically significant relationship detected in the entire correlation.

For 2011-2012, the value returned for the relationship between teachers' TVAAS data scores and teachers' classroom observation scores was 0.10, signifying, again, a non-existent or negligible statistical relationship between the two items. Similarly, the value returned for the relationship between teachers' student achievement data scores and teachers' classroom observation scores was 0.14, slightly more significant than that for the TVAAS-observation correlation, but still either non-existent or negligible in terms of statistical significance. In 2011-2012, the value returned for the relationship between value-added scores and student achievement scores was more significant than the other correlations this year but was statistically insignificant, compared with the TVAAS-achievement correlation for 2012-2013. The value returned for this relationship was 0.18, meaning that the relationship was either non-existent or negligible. These results confirm that inter-rater reliability among the different component parts of the teacher evaluation framework used in Shelby County Schools pose a threat to the integrity of the system. The district should continue to think carefully about how it will approach the issue of rater drift and the wide rift between scores given to teachers for each of the different elements of the evaluation. This not only has significant effects on teacher perceptions of the effectiveness,

fairness and accuracy of the system and, consequently, on teacher trust and faith in the system, but also calls into question, in a deeper sense, whether or not the system provides a fair and accurate depiction of teacher performance.

#### Side Effects of Teacher Evaluation Policy: Teacher Morale and Motivation.

I considered the subject of this portion of the report significant enough to include it separately from the rest of the discussion of teacher focus group discussion results earlier in this report. In addition, because this issue involves discussion of a side effect of the district's teacher evaluation policy and does not contribute to the primary focus of this report, the direct analysis of the program's effectiveness, this discussion is most properly placed in its own subsection. During the April 2014 Shelby County Schools teacher focus group series, a portion of the group's discussions were devoted to the issue of teacher morale and motivation as they might be affected by the district's teacher evaluation policy. During the focus group discussions, teachers were asked to respond to the following statement: "*I know of other teachers who have been significantly discouraged from teaching or who have stopped teaching for the district BECAUSE of the performance evaluation system*" (Shelby County Schools Teacher Focus Group Series 2014). Of the teachers who participated in the focus group series, nearly 70% responded by saying that they either knew of one or a few teachers or that they knew of several teachers who had either become significantly demotivated or who had left the district because of the performance evaluation system. An additional 10% of teachers concluded that they "might" know of another teacher in this situation. Twenty percent either responded that they knew of no

other teacher in this situation (10%) or that they were unsure (10%). The discussion moved further into specifically *what* about the teacher evaluation framework made these teachers either become significantly discouraged or stop working for the district. Teachers responded to the following question: “*What caused this teacher or teachers to become significantly discouraged or stop teaching for the district?*” (Shelby County Schools Teacher Focus Group Series 2014). I grouped teacher responses into the following categories. Some teachers responded, noting that the teacher or teachers in question left after receiving “fair performance evaluation scores” (10%).<sup>14</sup> Others noted that the teachers they knew in these situations had received “unfair performance evaluation scores” (30%) or that they “received insufficient support and feedback from the district” (20%), and that these factors had caused the teachers in question to become demoralized or stop teaching for the district. A series of other possible responses allowed for teachers who sought to identify other factors as the cause or who were uncertain as to the specific cause of the drop in teacher morale. Some teachers responded that they were unsure of why the teacher or teachers they knew had become demoralized or stopped working for the

---

<sup>14</sup> In order to account for the possibility of potential bias, the question was worded carefully so as to not ask teachers about their opinions about their own employment situations or about the opinions of other teachers about the employment situations of those teachers. Instead, in this question, teachers were specifically asked whether they knew of a teacher or teachers who had become significantly discouraged or stopped working for the district precisely because of the performance evaluation system. Once they confirmed whether or not they knew any teachers in this situation, teachers proceeded to answer questions about what specifically caused the teacher to become demoralized or stop teaching. They were asked to reflect on this answer, however, from their own perspective instead of being asked to recount what the teacher or teachers in question felt about their own situation. The thought was that teachers might know of a situation in which a teacher became discouraged or left because of the performance evaluation system, but, because they were more detached from the situation, would be able to more accurately reflect on what occurred. Obviously this does not eliminate all bias. There is a great likelihood that the teachers discussed in this portion of the focus group were friends or acquaintances of the teachers involved in the discussion. It is also likely that teachers might be sympathetic to fellow teachers who either left or became discouraged as the result of performance evaluation results simply because they could easily find themselves in the same positions as those fellow teachers. Still, by removing the personal interest of teachers in the answer they are giving as much as possible and by asking them to reflect objectively on what occurred to another teacher, the possibility for bias in teacher responses is at least minimized. Thus, whenever the terms “fair” or “unfair” were used in the discussion, teachers were asked to ensure that these judgements were being made from their perspectives and not from the perspectives of the teachers at issue in the discussion.

district (13.3%) Others noted that there were “other reasons” involved in the teacher’s demoralization or departure from teaching (16.7%). As noted in the earlier question, 10% of teachers confirmed that they did not know of any teachers in this situation. I argue that these results pose significant threats to the overall quality of the body of teachers in the greater Memphis area teaching in public schools and, specifically, to the equitable distribution of effective teachers in the district to the schools who arguably need them most, schools with high-poverty, at-risk students. While this report offers only a truncated and preliminary assessment of this significant side effect of teacher evaluation policy which the district ought to consider carefully, much future research has yet to be done in terms of garnering more detailed information about how many teachers leave or experience reduced effectiveness as a result of low district morale and what specific aspects of teacher evaluations are contributing to this attitude. Further still, more research must be conducted to investigate the overall effects of these policies for teacher quality in school districts across the country.

While these issues still require future research, there is a significant amount of literature that already exists on the current challenges facing teachers, especially in high-poverty, high-need schools. Teachers in these schools already experience very challenging work environments. When considering the effects of teacher evaluations on teacher morale, it is important to take the already-existing challenges to teacher morale into account. An interesting perspective on challenges facing teachers in their work environment alone, not even considering additional challenges posed by teacher evaluation systems, has been taken up in a series of recent studies which focus not necessarily on the *effectiveness* of individual teachers in contributing to their students’ learning, but on the *equity* of the school and community distribution of these highly

effective teachers in the United States. This is certainly an important component of the debate over the problem of teacher quality in the United States. If there are stronger levels of teacher quality (however it is measured) in certain wealthier or well-endowed communities or schools and significantly lower levels of teacher quality in schools attended by poorer students or students of color, then, as some would argue, a significant problem with teacher quality in the United States, as some have argued, would still exist. Research conducted by the Center for American Progress in 2014 found that, “In an analysis of the newest data, we find that in some areas, poor students and students of color are far less likely than others to have expert teachers” (DeMonte and Hanna 2014). The report analyzed information in two states, Louisiana and Massachusetts, “two early adopters of new teacher evaluation systems” which provide comprehensive yearly ratings for each teacher, assessing his or her effectiveness in contributing to student learning (DeMonte and Hanna 2014). DeMonte and Hanna cited a growing “concern among policymakers and student advocates...that students in disadvantaged schools are less likely to have access to high-quality instruction than students in affluent schools” (2014). This concern, they state, is the basis for their study. In their report, they used data in the two states “based on new accountability measures” to “glean information about the distribution of teachers across school demographics” (DeMonte and Hanna 2014). The information, they argue, “confirms previous findings” (DeMonte and Hanna 2014). They state: “In many places, poor children and children of color are less likely to be taught by a highly effective teacher” (DeMonte and Hanna 2014). In the data which the researchers studied from Louisiana, students in the poorest schools were “40 percent” less likely to be taught by a highly effective teacher than students in Louisiana’s wealthiest schools (DeMonte and Hanna 2014). The

numbers are similar when student racial demographics are considered. Students attending schools with the lowest percentages of minority students were “38 percent more likely to have a highly effective teacher” than students attending schools with what the researchers classified as a “high minority enrollment” (DeMonte and Hanna 2014).

Other studies have confirmed the results identified by the Center for American Progress. In 2000, Linda Darling-Hammond and Laura Post authored a chapter in Richard Kahlenberg’s *A Notion at Risk: Preserving Public Education as an Engine for Social Mobility*. In this chapter, they argued, “Few Americans realize that the U.S. educational system is one of the most unequal in the industrialized world, and students routinely receive dramatically different learning opportunities based on their social status” (Darling-Hammond and Post 2000). The chapter contrasted the U.S. educational system with other “European and Asian nations” which utilized centralized funding systems (Darling-Hammond and Post 2000). According to Darling-Hammond and Post, “the wealthiest 10 percent of school districts in the United States spend nearly ten times more than the poorest 10 percent, and spending ratios of three to one are common within states” (Darling-Hammond and Post 2000). These reduced levels of funding, the researchers argue, signal deeper issues with the educational quality students experience in these schools in low-income communities. Darling-Hammond and Post posit that “poor and minority students are concentrated in the less well funded schools, most of them located in central cities and funded at levels substantially below those of neighboring suburban districts” (2000). These funding differences, they argue, “leave minority students with fewer and lower-quality books, curriculum materials, laboratories, and computers; significantly larger class sizes;...less access to

high-quality curriculum” and, of significance to this report, these students consistently experience “less qualified and experienced teachers” (Darling-Hammond and Post 2000).

The problems in this area are deep-seated, Darling-Hammond and Post argue. They present data showing, for example, the number of teachers who were unlicensed in their main teaching field at schools with differing percentages of low-income and minority students. Only about 4% of schools with a low-income student population of less than 5% contained teachers who were unlicensed in their main teaching field (Darling-Hammond and Post 2000). In contrast, schools where more than fifty percent of students were economically disadvantaged were more than three times as likely to hire such unlicensed teachers (Darling-Hammond and Post 2000). Over 15% of newly hired teachers who were unlicensed in their main teaching field were hired by these types of schools (Darling-Hammond and Post 2000). The numbers were even more dramatic when student racial demographics were considered. Only about 3% of the new hires at schools with only 1-10% minority students were unlicensed in their main teaching field (Darling-Hammond and Post 2000). In schools where over fifty percent of the students were racial minorities, this number was slightly above 15% (Darling-Hammond and Post 2000). Disparities were found between schools with low numbers of low-income students and schools with high numbers of low-income students in terms of their rates of hiring teachers with Master’s degrees (Darling-Hammond and Post 2000). Wealthier schools hired new teachers with Masters’ degrees at a rate of almost twenty-five percent, whereas this number was under 15% for poorer schools (Darling-Hammond and Post 2000).

These distinctions can be understood in terms of need, because there is a greater need for highly effective teachers in schools with large proportions of low-income and racial minority

students. What the research shows, however, is that, even though these schools can be classified as “high-need,” they are less likely to hire and employ highly qualified or highly effective teachers, as defined by either traditional teacher qualifications of education and experience and also by information gathered from student standardized test scores. This has been identified by some researchers as significant problem in the area of teacher quality. Darling-Hammond and Post state: “This era is developing an even more sharply bimodal teaching force than ever before. While some children are gaining access to teachers who are more qualified and well-prepared than in years past, a growing number of poor and minority children are being taught by teachers who are sorely unprepared for the task they face” (2000). This, they state, “poses the risk that we may see heightened inequality in opportunities to learn and in outcomes of schooling- with all the social dangers that implies- at the very time we most need to prepare all students more effectively for the greater challenges they face” (Darling-Hammond and Post 2000). It is critical for districts to consider the effects of hiring and recruitment efforts not only the total number of highly effective teachers, but also on the distribution of these teachers as well, ensuring that high-need schools and classrooms are served by effective teachers just as often, if not more so, than are wealthier communities and students. Darling-Hammond and Post identify this as one of the most significant challenges facing both the issue of teacher quality in public schools and also the reform of education policy in general. They state: “If the emerging reforms of schooling are to succeed, and if students are to have a fair shot at meeting the high standards states and districts are increasingly insisting they meet, teaching as an occupation must be able to recruit and retain able and well-prepared individuals for all classrooms, not just the most affluent” (Darling-Hammond and Post 2000).



In light of these results, the feedback offered by teachers in these focus group discussions ought to be taken seriously by the district. The challenges facing teachers in a primarily urban school district, characterized by significant inequity among students and high levels of poverty, are already tending to discourage teachers, especially those serving in schools which are not naturally optimal teaching environments. This is why it is often extremely difficult to find teachers to serve in high-need, high-poverty schools or to serve in hard-to-staff positions, such as teaching English as a second language or special education classes. While current research shows that U.S. school districts do not necessarily face a shortage of teachers, there is a significant body of literature which argues that there *is* a shortage of highly effective teachers, however that is measured, in schools with the poorest, most at-risk students (Baker and Smith 1997, Wayne 2000). As Heather Voke argues, in a chapter in Marge Scherer's *Keeping Good Teachers*, "Researchers...dispute the conventional wisdom that [a teacher] shortage exists because there are simply not enough qualified teachers to fill the number of vacant positions" (Voke 2003). To the contrary, Voke asserts, "If we consider only the number of qualified candidates and the number of job openings, there is an overall surplus of trained people (2003, citing Darling-Hammond 2001, NASBE 1998). Voke argues that there *is* a shortage of teachers, but that the shortage "lies in the distribution" and not in the overall number of teachers (2003). She argues that: "There are not enough teachers who are both qualified and willing to teach in urban and rural schools, particularly in those serving low-income students or students of color" (Voke 2003). She adds, "There is also a shortage in certain geographic regions of the country, and there are not enough qualified individuals in particular specialties, such as special education, bilingual education, and the sciences" (Voke 2003, citing Bradley 1999, NASBE 1998). Voke adds that

there is also a significant issue involving “high rates of teacher turnover,” even though there is not a traditional shortage in terms of overall number of teachers (2003, citing Ingersoll 2000).

As stated, 80% of teacher participants claim to know of a teacher or teachers who have either become significantly discouraged or who have left teaching for the district and 50% confirm that these teachers have left as a result of some fault in the evaluation system, either due to unfair performance evaluation scores or due to the failure of the district to provide effective feedback and support to help teachers improve. From these results, it seems as though the current teacher evaluation framework may be exacerbating the already-difficult challenges that threaten teacher morale and motivation, especially at high-need schools and in hard-to-staff areas and positions. Teachers increasingly point to the Shelby County Schools’ teacher evaluation framework as time-consuming and demanding and, at times, as more focused on assigning ratings to teachers rather than helping them improve. While considering how it will continue to improve and implement its teacher effectiveness framework, the district ought to pay careful attention to the potentially harmful effects of the current structure and implementation of its teacher evaluation policies on teacher morale and motivation. Given the already-proven existence of the difficulty of attracting high quality teachers to high-need school, the district ought to consider how it will continue to attract and retain highly effective teachers, especially in its poorest and most understaffed schools in light of teacher perceptions and reactions to its teacher evaluation regimen. If teachers who are currently invested in the essential work of urban education and who work specifically with high-need and at-risk student populations are being turned away or are experiencing significant declines in effectiveness as a result of low morale, the district must consider how it will re-energize or replace these teachers, should they decide to

leave the district for private or municipal school district teaching positions, or decide to leave its high-need schools for schools populated by more affluent students and fewer students of color. Teacher morale is an essential factor in the health of any school district and, as evidenced by teacher responses in focus group discussions, the teacher evaluation program implemented by Shelby County Schools has the potential to seriously harm this morale. As one teacher stated in the course of focus group discussions: “Morale is at an all-time low. I have been teaching for 17 years, and I have never seen such depressed, downtrodden teachers in my entire life” (Shelby County Schools Teacher Focus Group Series 2014). Another teacher responded during this portion of the focus group discussions by noting that the teacher was aware of other teachers who had stopped teaching in the district for reasons related to the evaluation system. When asked why these teachers left, the teacher responded:

They left or retired because of the ridiculous demands placed on teachers- particularly [that] of the evaluation system... We went from being evaluated once every 5 years to being evaluated 3-6 times a year. That is crazy. I don't mind an administrator walking in my room any day of the week, but the lengthy conferences, paperwork, and planning that is necessary for every evaluation is [far] too time-consuming. The stress caused by having to earn all 5's is enough to give us ulcers. (Shelby County Schools Teacher Focus Group Series 2014).

In order to halt any further damage from being done to teacher morale and motivation in the district, which is so essential to the success of students in the greater Memphis area, specifically those in high-need, high-poverty schools, the district ought to carefully study and evaluate the effect the teacher evaluation system has on teacher morale.

## Conclusion

The results of this study cannot be oversimplified or reduced to a simple assessment of the effectiveness of the teacher evaluation policy implemented by Shelby County Schools. As has already been established, teaching and learning is a complex process and the research which attempts to ascertain how the process works or what factors most significantly influence it has almost undoubtedly run into difficulty attempting to make generalizations from specific case studies of schools, teachers, and groups of students. Thus, it should not come as a surprise that attempting to evaluate and rate the work of teachers on a scale ranging from one to five is not an easy or straightforward process either. If this report shows anything, it is that the ongoing work of developing a comprehensive way to both hold public education providers accountable for the quality of their work to a public ever more attentive to public education is still, most certainly, a work in progress. As advanced as data collection and usage have become in public education systems over the past few decades, they are still far from precision. It must hold then, that decisions based on such information ought always to be qualified and understood as confined by their limitations. The district's attentiveness to the importance of teachers in affecting student achievement is most certainly appropriate as is the district's conviction that it must take action to attempt to right the inequities it sees in student performance across both city and county limits. Issues such as the influence of state and federal policymaking, the growing importance of private providers of public education, the charter schools, the contributions of private organizations and foundations in funneling best practices ideas and research into the hands of district policymakers, significantly increased public attention to education reform all serve to make these foci and

convictions more complicated. Local concerns such as those brought about by the school merger or by the unique demographic characteristics of the population served by Shelby County Schools make approaching these issues even more difficult.

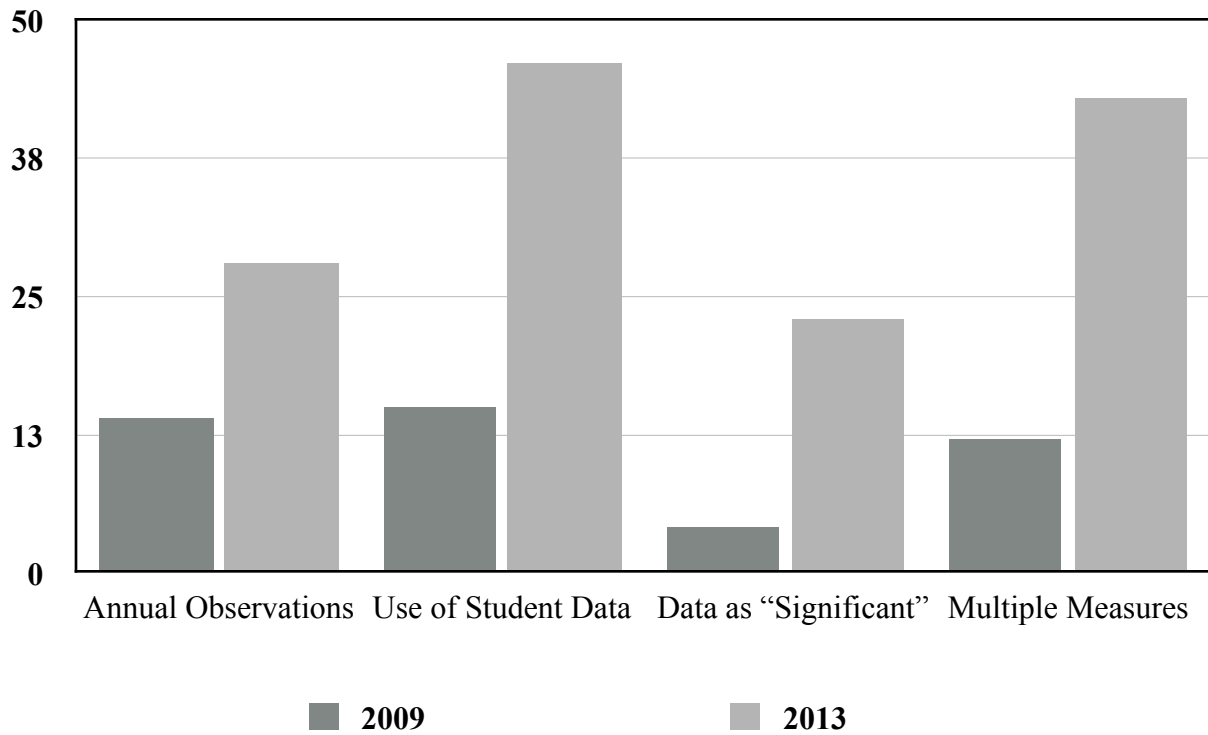
The district has made significant strides toward providing a strong definition of effective teaching and learning. This definition is research-based and has been developed and its design has been overseen by teachers and practitioners working in the greater Memphis area. The district has developed a highly detailed and complex mechanism by which to implement and put into working practice this definition in the form of the teacher evaluation system analyzed in this report. On top of this, the district has sought to use the performance evaluation system not only to catalog and document teacher performance and to identify weak performers, but also to provide in real time, feedback and support needed to help teachers improve their practice throughout the school year. It is very difficult to fault the district on the grounds of its intent, as it has consistently pressed for teacher evaluations to be data-driven and to be used to improve students achievement, by helping teachers improve their practice. Where the difficulty and the controversy lie, however, is in the district's ongoing implementation of its objectives and in its keeping up its commitment to involve teachers and be attentive to their concerns about the system. There is significant negativity in terms of teacher perceptions of the system which the district ought to study and pay considerable attention to as it has done in the past. These are concentrated in teacher doubts about the fairness and accuracy of performance evaluation scores, both in the way they are obtained as well as in the way they are used for district-wide personnel decision-making. Teachers also have weak trust in the use of certain components for the work of performance evaluation at all. These include teacher perceptions of the use of TVAAS data in an

evaluation context and the use by the district of student perception surveys to rate teachers. Even beyond this, the district ought to pay attention to teacher arguments regarding the issue of their being held accountable for student performance to the degree that they are being either punished or rewarded for significant external factors which they cannot control. As Carolyn Heinrich has suggested, this has the potential to demotivate and demoralize teachers, causing significant problems for the existing teacher corps in Memphis and Shelby County. The results of the focus group and data analysis research confirm, in their own limited way, the need for the district to listen to these arguments.

The district has attempted, as Robert Hutchins argued in the 1930's, to press the issue of education reform beyond the provision of public education and into an investigation and assessment of the quality of the services its teachers provide in their classrooms. This stems from its desire to *actually* achieve its mission of improving student achievement, particularly for its students who face the highest risk and who possess the highest needs. This report shows, however, the need to approach such a mission carefully and thoughtfully. In recent years, the district has tended to oversimplify the relationship between a teacher's effect and a student's performance, which has led it, at least in the eyes of its critics and some of its teachers, to blame teachers for poor performance. As past studies have shown and as this study confirms, there are a host of factors beyond the control, not only of the school district or school, but also of the individual teacher, for which teachers are now being held accountable. This has already begun to have early effects on teacher morale, motivation and, consequently, teacher attrition. The district must continue to cautiously and conscientiously approach the subject of teacher evaluations, always being ready to listen attentively to the advice of those who serve its students in the

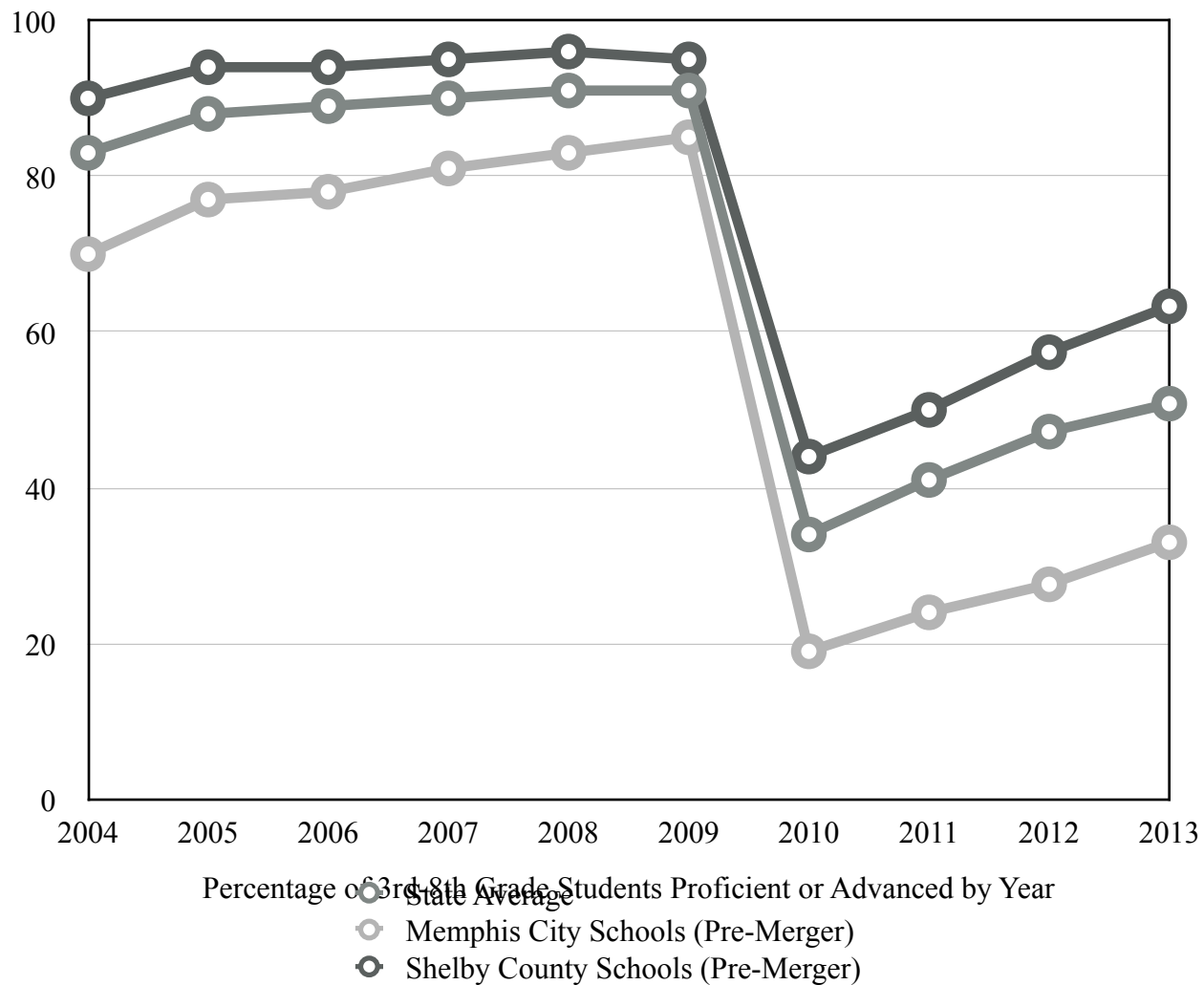
classrooms and who are daily subject to its evaluation requirements. It ought to place considerable emphasis on securing the buy-in and support of teachers. In some areas, such as its professional development system and when it comes to classroom observations by principals, teachers are generally supportive of the district. The district must continue, however, to seek authentic and frequent teacher feedback regarding the effectiveness of its performance evaluation system and to continually ensure that the assumptions on which its system relies are research-based and sustain a healthy, vibrant, challenging and supportive environment for teaching and learning which will continue to both motivate and improve current teachers as well as attract potential candidates. In this way, the district can secure the opportunity to move forward and generate a lasting reform of the system in order to realistically improve educational offerings for all students in the greater Memphis area.

## TABLES AND GRAPHS

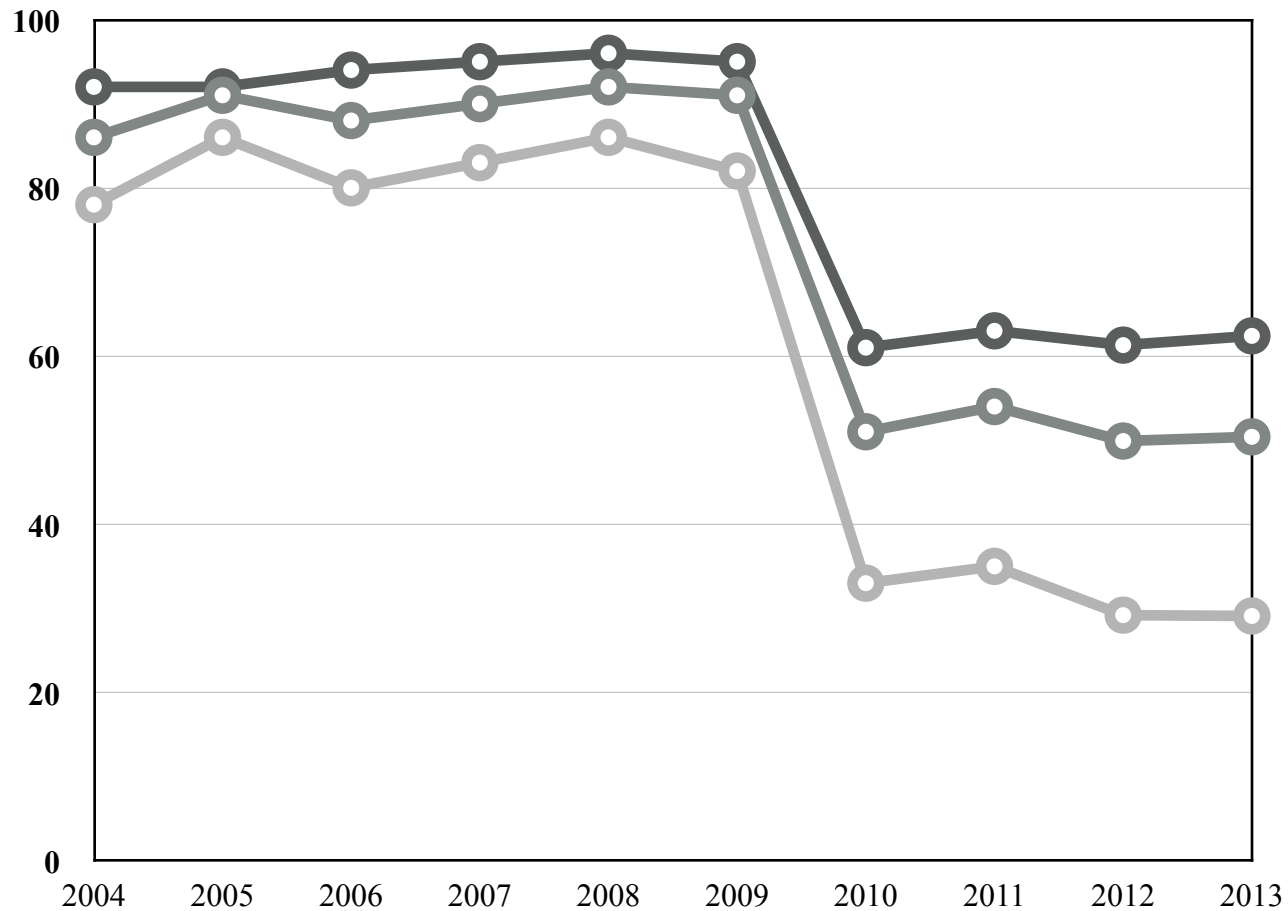
**Table 1. 2009-2013 State Teacher Evaluation Legislation**



**Table 2. TCAP Criterion-Referenced Academic Achievement: Math, Grades 3-8**



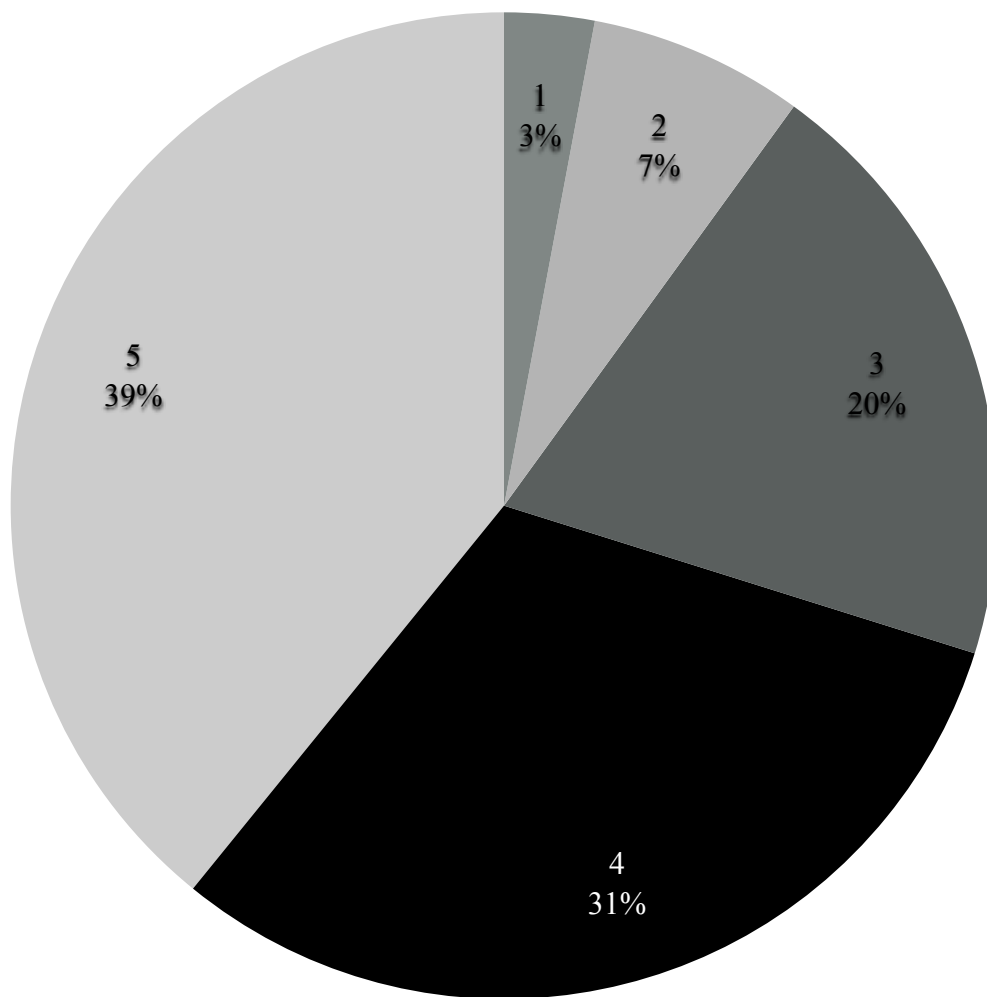
**Table 3. TCAP Criterion-Referenced Academic Achievement: Reading/  
Language Arts, Grades 3-8**



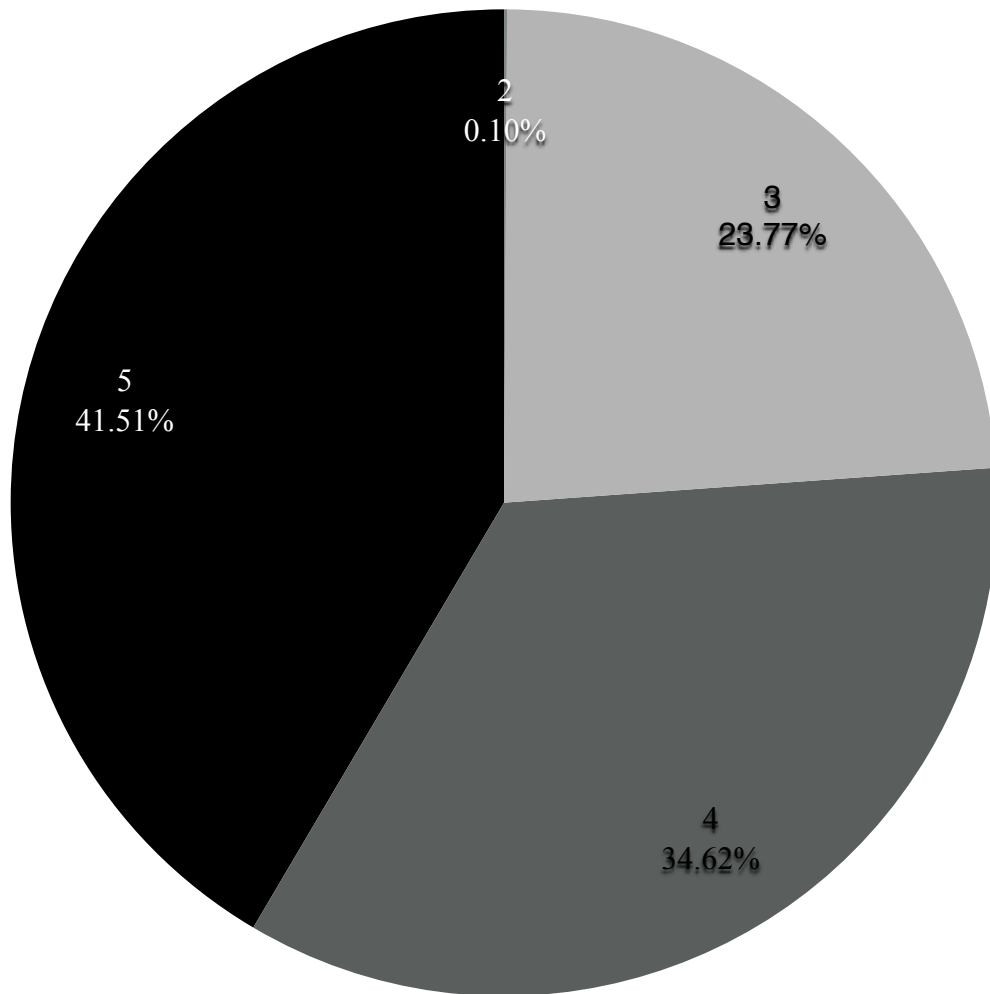
**Percentage of 3rd-8th Grade Students Proficient or Advanced by Year**

- State Average
- Memphis City Schools (Pre-Merger)
- Shelby County Schools (Pre-Merger)

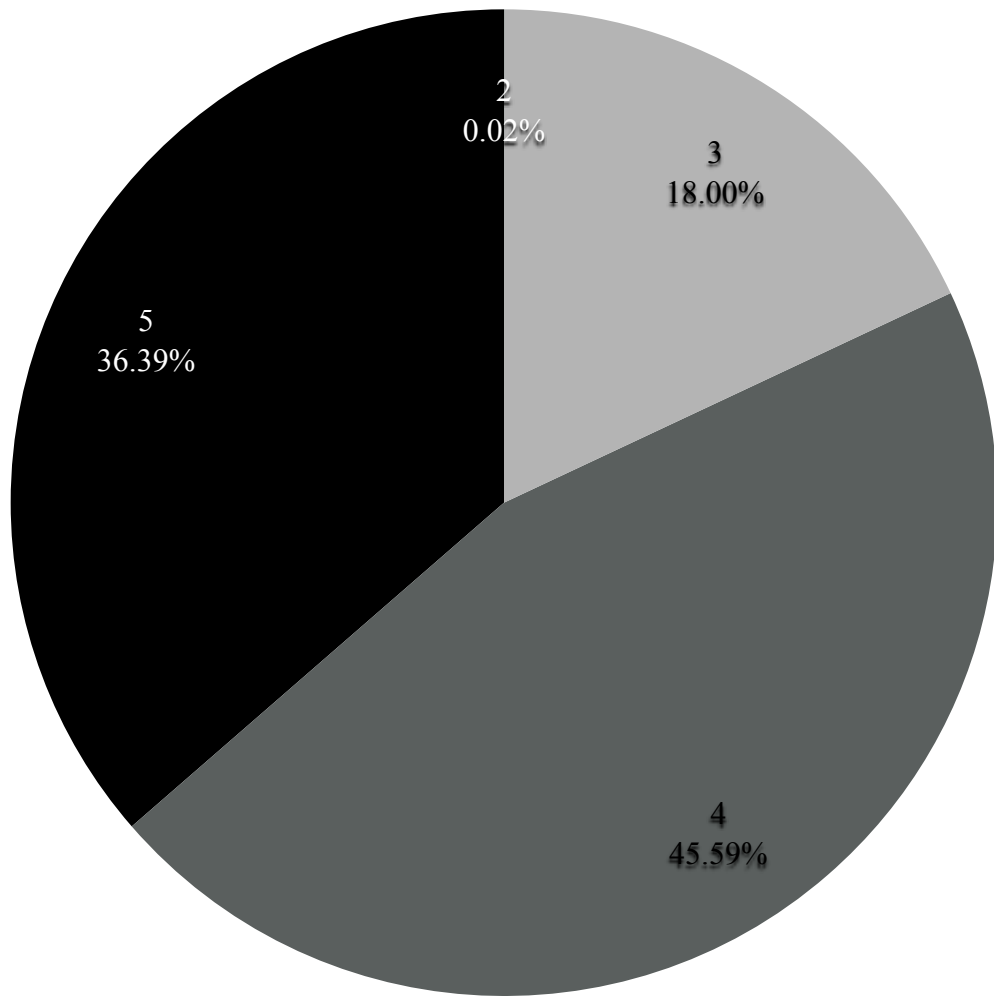
**Table 4. 2013 Teacher Evaluation Summative Scores for Memphis City Schools**



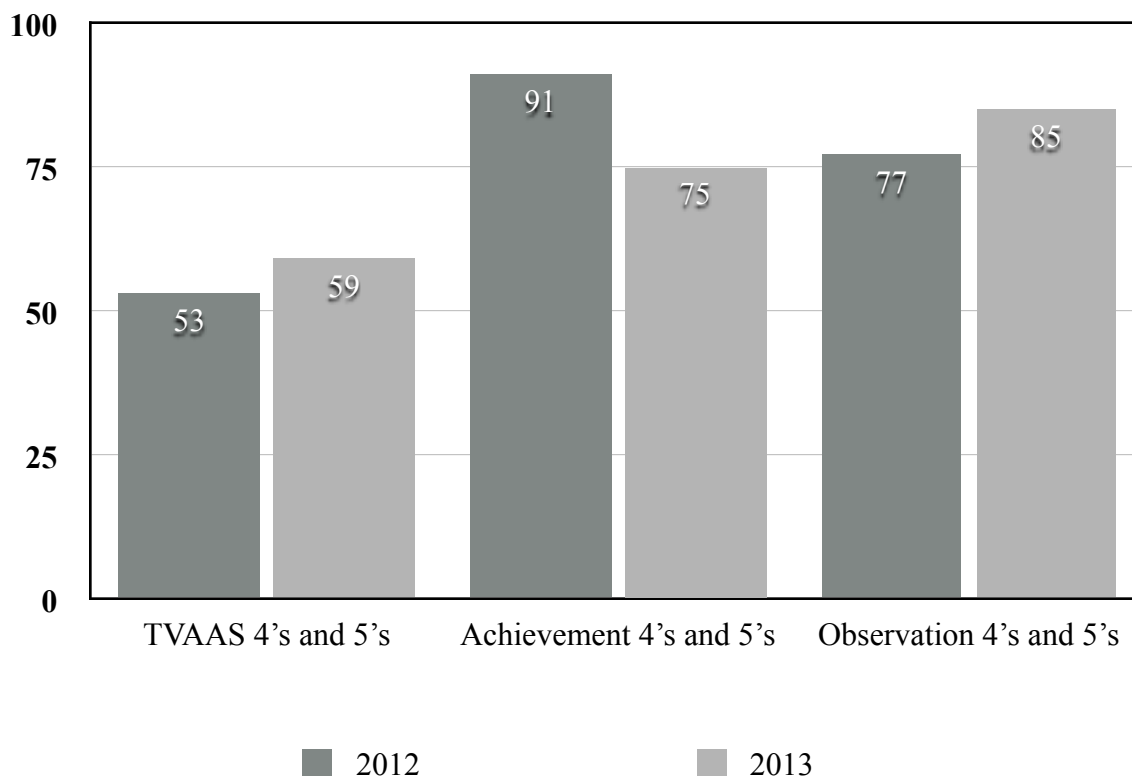
**Table 5. 2012 Teacher Evaluation Summative Scores for Memphis and Shelby County**



**Table 6. 2013-2014 Partial Teacher Classroom Observation Scores (Rated by Principals)**



**Table 7. Percentage of Teachers Rated as 4 or 5 on Classroom Observation, TVAAS and Student Achievement Data Scores in 2011-2012 and 2012-2013**



## REFERENCES

- Ammons, David N. 1992. "Productivity Barriers in the Public Sector." Ed. Marc Holzer. *Public Productivity Handbook*, 117-136. New York, NY: Marcel Dekker, Inc.
- Ammons, David N. 1995. "Overcoming the Inadequacies of Performance Measurement in Local Government: The Case of Libraries and Leisure Services." *Public Administration Review* 55.1: 37-47.
- Amrein-Beardsley, Audrey, and Clarin Collins. 2012. "The SAS Education Value-Added Assessment System (SAS EVAAS) in the Houston Independent School District (HISD): Intended and Unintended Consequences." *Education Policy Analysis* 20.12.
- Arthur, E. F. 1987. "The Ethics of Corporate Governance." *Journal of Business Ethics* 6: 59-70.
- Bailey, Clay, "Bartlett residents discuss creating separate school district," *The Commercial Appeal*, November 15, 2011.
- Bailey, Clay, "Suburban municipal residents feel disconnect, lack of trust with unified school system," *The Commercial Appeal*, February 1, 2012.
- Bailey, Clay, "Six Memphis suburbs give OK to seek Aug. 2 municipal school referendums," *The Commercial Appeal*, May 29, 2012.
- Bailey, Clay, "Shelby schools merger pace troubles judge, prompts call for master to oversee process," *The Commercial Appeal*, February 25, 2013.
- Bailey, Clay, "Consultant confident Germantown can open schools next summer," *The Commercial Appeal*, September 13, 2013.
- Bailey, Clay, Tom Charlier, Sara Patterson, and Cindy Wolff, "Shelby County suburbs approve starting schools," August 2, 2012.
- Bailey, Clay, Daniel Connolly, and Michael Kelley, "Commissioners seek to block August municipal school district referendum," June 26, 2012.
- Bailey, Clay and Richard Locker, "Shelby County suburban leaders go to Nashville to push for 'local control' of schools," *The Commercial Appeal*, February 5, 2013.
- Bakeman, Jessica, "State announces high teacher scores, hopes union fears are calmed," *Capital New York*, October 22, 2013.
- Baker, Bruce D. 2012. "The Toxic Trifecta, Bad Measurement & Evolving Teacher Evaluation

- Policies." *National Education Policy Center*.
- Baker, D. P., and T. Smith. 1997. "Teacher turnover and teacher quality: Refocusing the issue." *Teachers College Record* 99.1: 29–35.
- Baker-Doyle, Kira J. 2011. *The networked teacher: How new teachers build social networks for professional support*. New York/London: Teachers College Press.
- Balfanz, Robert and Nettie Legters, Thomas C. West and Lisa M. Weber. 2007. "Are NCLB's Measures, Incentives, and Improvement Strategies the Right Ones for the Nation's Low-Performing High Schools?" *American Educational Research Journal* 44.3: 559-593.
- Ballou, D. 2002. "Sizing up test scores." *Education Next* (Summer 2002): 10-15.
- Banner, D. K., and R.A. Cooke. 1984. "Ethical Dilemmas in Performance Appraisal." *Journal of Business Ethics* 3: 327-333.
- Bartlett, S. 2000. "The Development of Teacher Appraisal: A Recent History." *British Journal of Educational Studies*. 48.1: 24-37.
- Barrett, R.S. 1966. *Performance Rating*. Chicago, IL: Science Research Associates Inc.
- Barton, A.C., Drake, C., Perez, J.G., St. Louis, K., & George, M. 2004. "Ecologies of Parental Engagement in Urban Education." *Educational Researcher* 33: 3-12.
- Batley, Richard and George A. Larbi. 2004. *The changing role of government: The reform of public services in developing countries*. London: MacMillan.
- Beary, Rodney. 1990. "If You Really Care, Tell Them Like it is." *PIMA Magazine* 72.8: 19.
- Behn, Robert D. 2001. *Rethinking Democratic Accountability*. Washington, DC: Brookings Institution.
- Behn, Robert D. 2003. "Why Measure Performance? Different Purposes Require Different Measures." *Public Administration Review* 63.5: 586-606.
- Bernardin, H. J. 1981. "Rating training strategies: An integrative model." Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.
- Bernardin, H. J., and R.W. Beatty. 1984. *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent.
- Bernardin, H. John, and Michael Wiatrowski. 2013. "Performance appraisal." *Psychology and*



*Policing: 257.*

- Bertelli, Anthony M. and Peter John. 2010. "Government Checking Government: How Performance Measures Expand Distributive Politics." *The Journal of Politics* 72.2: 545-558.
- Besterfield, Dale H., Carole Besterfield-Michna, Glen Besterfield, and Mary Besterfield-Sacre. 2003. *Total Quality Management*, 3rd Ed. New Jersey: Prentice Hall.
- "Empowering Effective Teachers." 2000. *Bill and Melinda Gates Foundation*.
- The Bill and Melinda Gates Foundation (Anonymous Source). Personal Interview by Landon Webber, April 23, 2014.
- Blanchard, Ken. 1996. "Giving and Receiving Feedback." *Manage* 48.1: 31.
- Blodgett, Terrell, and Gerald Newfarmer. 1996. "Performance Measurement: (Arguably) The Hottest Topic in Government Today." *Public Management* 6.
- Bornfreund, Laura. 2013. *How Federal Policy Can Promote Better Prepared Teachers and School Leaders*. Education Policy Program Brief. Washington, D.C.: The New American Foundation.
- Bouckaert Geert and B. Guy Peters. 2002. "Performance Measurement and Management: The Achilles' Heel in Administrative Modernisation." *Public Performance & Management Review* 25.4: 359-362.
- Bozeman, Barry. 1987. *All Organizations are Public: Bridging Public and Private Organizational Theory*. San Francisco: Jossey-Bass.
- Bradley, A. 1999. "States' uneven teacher supply complicates staffing of schools." *Education Week* 18.26: 1, 10-11.
- Branston, John, "Test Scores in for Memphis and Shelby County Schools," *Memphis Flyer*, July 26, 2012.
- Brief, A. P., R.S. Schuler and M. Van Sell. 1981. *Managing job stress*. Boston: Little, Brown.
- Brown, W. 2000. "Reporting NAEP achievement levels: An analysis of policy and external reviews." Eds. M. Bourque and S. Byrd. *Student performance standards on the National Assessment of Educational Progress: Affirmation and improvements*, 11-40. Washington, D.C.: National Assessment Governing Board.

- Brumback, Gary B. 2003. "Blending "we/me" in performance management." *Team Performance Management* 9.7/8: 167-173.
- Brumback, Gary B. 2010. "Performance Management: Putting Research into Action." *Personnel Psychology* 63.2: 501-507.
- Brumback, G.B. and J.W. Vincent. 1970. "Jobs and Appraisal of Performance." *Personnel Administration* 33: 26-30.
- Buchholz, R. A. 1982. *Business environment and public policy*. Englewood Cliffs, NJ: Prentice Hall.
- Buchholz, R. A. 1985. *Essentials of public policy for management*. Englewood Cliffs, NJ: Prentice Hall.
- Buchholz, R. A. 1989. *Business ethics*. Englewood Cliffs, NJ: Prentice Hall.
- Buser, Lawrence, "Shelby County Schools file federal suit," *The Commercial Appeal*, February 11, 2011.
- Buser, Lawrence, "At judge's request, Shelby County Commission agrees to delay appointment of new school board," *The Commercial Appeal*, March 24, 2011.
- Buser, Lawrence, "Judge tells parties to agree on unified Shelby County Schools board or he'll decide," *The Commercial Appeal*, April 24, 2011.
- Butrymowicz, Sarah and Sarah Garland. 2012. "How New York City's value-added model compares to what other districts, states are doing." *The Hechinger Report*, March 1, 2012. [http://hechingerreport.org/content/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing\\_7757/](http://hechingerreport.org/content/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing_7757/) (April 25, 2013).
- Callanan, Mark. 2010. "Who designs local government performance measurement systems?" *International Journal of Public Sector Performance Management* 1.4: 346-359.
- Camilleri, Emanuel and Beatrice I. J. M. Van Der Heijden. 2007. "Organizational Commitment, Public Service Motivation, and Performance within the Public Sector." *Public Performance and Management Review* 31.2: 241-274.
- Campbell, Donald J., and Cynthia Lee. 1988. "Self-Appraisal in Performance Evaluation: Development versus Evaluation." *The Academy of Management Review* 13.2: 302-314.
- Canon, Gabrielle, "Memphis and Shelby County Schools Merger Prompts Battle Over Politics, Race and Money," *The Huffington Post*, May 25, 2011. <http://www.huffingtonpost.com/>

- 2011/03/15/memphis-shelby-county-schools-merger\_n\_836333.html (April 3, 2014).
- Carroll, S. J., C. Lee, M.S. Taylor, & D.J. Gillen. 1985. "Task uncertainty as a moderator of the feedback-performance relationship." Paper presented at the annual meeting of the Southern Management Association, Orlando, FL.
- Chan, Hon S. and Jie Gao. 2009. "Putting the Cart before the Horse: Accountability or Performance?" *Australian Journal of Public Administration Special Issue: Accountability and Public Governance in Asia* 68.Issue Supplement, s1: S51-S61.
- Cherrington, DJ. and J.O. Cherrington. 1974. "Participation, Performance and Appraisal." *Business Horizons* 17.
- Chetty, Raj and John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach and Danny Yagan. 2010. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." Cambridge, Massachusetts: National Bureau of Economic Research. *NBER Working Paper Series*, Working Paper 16381.
- Clotfelter, C. T., H.F. Ladd, and J.L. Vigdor. 2007. "Teacher-student matching and the assessment of teacher effectiveness." *NBER Working Paper Series*, Working Paper 11936. Cambridge, MA: The National Bureau of Economic Research (NBER).
- Cohen, David K. and Deborah Loewenberg Ball. 1990. "Policy and Practice: An Overview." *Educational Evaluation and Policy Analysis* 12.3: 233-239.
- "A Tale of the Tape: Comparing Memphis City Schools and Shelby County Schools," *The Commercial Appeal*, 2011. [web.commercialappeal.com/school\\_timeline/tale-of-the-tape.html](http://web.commercialappeal.com/school_timeline/tale-of-the-tape.html) (April 15, 2013).
- "Germantown, Collierville start process to form their own school districts," *The Commercial Appeal*, January 23, 2012.
- Congressional Budget Office. 1986. *Trends in Educational Achievement*. Washington, D.C.: CBO.
- Congressional Budget Office. 1987. *Educational Achievement: Explanations and Implications of Recent Trends*. Washington, D.C.: CBO.
- Connolly, Daniel, "Shelby County Commission finalizes vote for 25-member unified school board," *The Commercial Appeal*, February 28, 2011.
- Cornett, Lynn M. 1995. "Lessons from 10 Years of Teacher Improvement Reforms." *School Reform: What We've Learned*. 52.5: 26-30.

- Corporate Industrial Psychologist (Anonymous Source). Personal Interview by Landon Webber and Kemmashela Smith, July 12, 2013.
- Cutt, James. 1982. "Accountability, Efficiency, and the 'Bottom Line' in Non-Profit Organizations." *Canadian Public Administration* 25: 311-331.
- Darling-Hammond, Linda. 2000. *Solving the dilemmas of teacher supply, demand, and standards*. New York: National Commission on Education and America's Future.
- Darling-Hammond, Linda. 2001. "The challenge of staffing our schools." *Educational Leadership* 58.8: 12-17.
- Darling-Hammond, Linda. 2013. *Getting Teacher Evaluation Right: What Really Matters for Effectiveness and Improvement*. New York: Teachers College.
- Darling-Hammond, Linda and D.L. Ball. 1998. *Teaching for high standards: What policymakers need to know and be able to do*. Philadelphia, PA: National Commission on Teaching and America's Future and the Consortium for Policy Research in Education.
- Darling-Hammond, Linda, Audrey Amrein-Beardsley, Edward Haertel, and Jesse Rothstein. 2012. "Evaluating teacher evaluation." *The Phi Delta Kappan* 93.6 (March): 8-15.
- Darling-Hammond, Linda and Laura Post. 2000. "Chapter 5: Inequality in Teaching and Schooling- Supporting High-Quality Teaching and Leadership in Low-Income Schools." *A Notion at Risk: Preserving Public Education as an Engine for Social Mobility*. Ed. Richard D. Kahlenberg. New York, NY: The Century Foundation/Twentieth Century Fund, Inc.
- Darling-Hammond, Linda, and Gary Sykes. 1999. *Teaching as the learning profession: Handbook of policy and practice*. San Francisco, CA: Jossey-Bass Publishers.
- Davis, P.A. 1974. "Do They Agree with Your Appraisals of Their Performance?" *Supervisory Management* 1: 19.
- Davis, Tom and Michael Landa. 1999. "Pat or Slap? Do Appraisals Work?" *CMA Management* 73.2: 24-26.
- Deal, Terrence E. and Allan A. Kennedy. 1982. *Corporate Cultures: The Rights and Rituals of Corporate Life*. Reading, Mass.: Addison Wesley.
- Decotiis, Thomas, and Andrew Petit. 1978. "The Performance Appraisal Process: A Model and Some Testable Propositions." *The Academy of Management Review* 3.3: 635-646.

- Deets, N.R. and D.T. Tyler. 1986. "How Xerox Improved its Performance Appraisals." *Personnel Journal* 65.4: 50-52.
- Deming, W.E. 1986. *Out of the Crisis*. Cambridge, MA: Center for Advanced Engineering Study, Massachusetts Institute of Technology.
- DeMonte, Jenny and Robert Hanna. 2014. *Looking at the Best Teachers and Who They Teach: Poor Students and Students of Color are Less Likely to Get Highly Effective Teaching*. Washington, D.C.: Center for American Progress.
- Deslandes, R. & Bertrand, R. N.P. "Motivation of Parent Involvement in Secondary-Level Schooling." *Journal of Educational Research* 98: 164-175.
- Dillon, Sam, "Merger of Memphis and County School Districts Revives Race and Class Challenges," *The New York Times*, November 5, 2011.
- Dries, Bill, "Hopson: 300 Employees to Lose Jobs," *The Memphis Daily News*, June 26, 2013.
- Duncan, Arne, "The Tennessee Story," *The Huffington Post Education Blog*, July 23, 2012. The Huffington Post, 23 July 2012. [http://www.huffingtonpost.com/arne-duncan/the-tennessee-story\\_b\\_1695467.html](http://www.huffingtonpost.com/arne-duncan/the-tennessee-story_b_1695467.html) (April 25, 2013).
- Dunnette, M. D. and W.C. Borman. 1979. "Personnel selection and classification systems." *Annual Review of Psychology* 30: 477-525.
- Dwivedi, R.S. 2001. "Developing a Culture of High Performance." *Indian Journal of Industrial Relations* 37.1: 31-57.
- Editorial Board of the New York Times, "Why Other Countries Teach Better: Three Reasons Students Do Better Overseas," *The New York Times*, December 17, 2013.
- "Teacher Quality." July 8, 2011. *Education Week*.
- Elmore, R. F., C. H. Abelman, et al. 1996. "The New Accountability in State Education Reform." Ed. Helen Ladd. *Holding Schools Accountable*, 65-98. Washington, D.C.: The Brookings Institution.
- Epstein, J.L. 1991. "Effects on student achievement of teachers' practices of parent involvement." Ed. S.B. Silvern. *Advances in reading/language research: Literacy through family, community, and school interaction* (Vol. 5, pp. 261-276). Greenwich, CT: JAI Press.

- Epstein, J.L., and Van Voorhis, F.L. 2001. "More than minutes: Teachers' roles in designing homework." *Educational Psychologist* 36, 181-193.
- Ewoh, Andrew I.E. 2011. "Performance Measurement in an Era of New Public Management." *Journal of Emerging Knowledge on Emerging Markets* 3.1: 8.
- Feldman, J. M. 1981. "Beyond attribution theory: Cognitive processes in performance appraisals." *Journal of Applied Psychology* 66: 127-148.
- Finch, Maida A. 2012. "Precursors to Policy Innovation: How Tennessee Entered Race to the Top." *Peabody Journal of Education* 87.5: 576-592.
- Fuller, Bruce and Joseph Wright, Kathryn Gesicki, and Erin Kang. 2007. "Gauging Growth: How to Judge No Child Left Behind?" *Educational Researcher* 36.5: 268-278.
- Garland, Sarah, "Tennessee Evaluation Systems Have Rough Road Ahead," *The Huffington Post Education Blog*, February 7, 2012. [http://www.huffingtonpost.com/2012/02/07/tennessee-teacher-evaluat\\_n\\_1260790.html](http://www.huffingtonpost.com/2012/02/07/tennessee-teacher-evaluat_n_1260790.html) (April 25, 2013).
- Garlington, Lela, "Voters say 'yes' to separate school districts in Memphis suburbs," *The Commercial Appeal*, July 16, 2013.
- Gellerman, S. W. 1976. *The Management of Human Resources*. Hinsdale, IL: Dryden Press.
- Ghorpade, Jai, Milton M. Chen and Joseph Caggiano. 1995. "Creating Quality-Driven Performance Appraisal Systems [and Executive Commentary]." *The Academy of Management Executive* (1993-2005) 9.1: 32-41.
- Glickman, A.S. 1955. "Effects of Negatively Skewed Ratings on Motivation of the Rated." *Personnel Psychology* 8.
- Goertz, Margaret, and Mark Duffy. 2001. *Assessment and Accountability Across the 50 States*. Philadelphia, PA: The Consortium for Policy Research in Education.
- Gold, Kenneth A. 1982. "Managing for Success: A Comparison of the Private and Public Sectors." *Public Administration Review* 412: 568-575.
- Grace, J. Peter. 1984. *War on Waste, President's Private Sector Survey on Cost Control*. New York: Macmillan.
- Green, Preston C., Bruce D. Baker, and Joseph Oluwole. 2012. "The Legal and Policy Implications of Value-Added Teacher Assessment Policies." *Brigham Young University Education & Law Journal* 1.

- Greenstone, Michael and Adam Looney. 2011. "Trends: Reduced Earnings for Men in America." Washington, D.C.: Brookings Institution. *Milken Institute Review*, Third Quarter 2011.
- Griffith, James. 2003. "Schools as Organizational Models: Implications for Examining School Effectiveness." *The Elementary School Journal* 104:1: 29-47.
- Grote, Richard C. 1996. *The Complete Guide to Performance Appraisal*. New York: AMACOM.
- Halachmi, Arie. 2005a. "Performance measurement is only one way of managing performance." *International Journal of Productivity and Performance Management* 54.7: 502-516.
- Halachmi, Arie. 2005b. "Performance measurement: test the water before you dive in." *International Review of Administrative Sciences* 71: 255-266.
- Hanushek, E. A. 1992. "The trade-off between child quantity and quality." *Journal of Political Economy* 100: 85-117.
- Hanushek, E. A. 2009. "Teacher deselection." Eds. D. Goldhaber and J. Hannaway. *Creating a new teaching profession*. Washington, DC: Urban Institute Press.
- Hanushek, Eric A., John F. Kain and Steven G. Rivkin. 1998. "Teachers, Schools, and Academic Achievement." *National Bureau of Economic Research Working Paper Series*, Working Paper 6691.
- Hanushek, E. A., J. Kain, and S. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39: 326-354.
- Hanushek, Eric A. and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *The American Economic Review* 100.2 (May): 267-271.
- Hanushek, E. A. and S. Rivkin. 2007. "Pay, working conditions and teaching quality." *The Future of Children* 17.1: 69-86.
- Harbison, R. W., and Eric A. Hanushek. 1992. *Educational performance for the poor: Lessons from rural northeast Brazil*. Oxford, England: Oxford University Press.
- Hardy, Kevin, "Tennessee's New Evaluation System Takes Shape," *Chattanooga Times Free Press*, November 22, 2011.
- Harris, D. N., and T.R. Sass. 2007. "Teacher training, teacher quality, and student achievement." Working Paper No. 3. Washington, DC: National Center for Analysis of Longitudinal

Data in Education Research.

- Hatry, Harry P. 1980. "Performance Measurement Principles and Techniques: An Overview for Local Government." *Public Productivity Review* IV: 312-339.
- Heathfield, Susan. 2000. "Performance Appraisals Don't Work." *Human Resources*.
- Heinrich, Carolyn J. 2002. "Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness." *Public Administration Review* 62.6: 712-725.
- Heitin, Liana, "Take It From Tennessee: Lessons on Teacher Evaluation," *Education Week*, November 27, 2012.
- Heneman, Herbert G., and Anthony T. Milanowski. 2004. "Alignment of Human Resource Practices and Teacher Performance Competency." *Peabody Journal of Education* 79.4: 108-125.
- Henry, H.W. 1962. "Performance Evaluation of Scientists: Current Methods." *Management of Personnel Quarterly* 13.1.
- Herold, D. M., and M.M. Greller. 1977. "Feedback: The definition of a construct." *Academy of Management Journal* 20: 142-147.
- Hickman, Craig R. and Michael A. Silva. 1984. *Creating Excellence: Managing Corporate Culture, Strategy and Change in the New Age*. New York: New American Library.
- Hightower, Amy M., Rachel C. Delgado, Sterling C. Lloyd, Rebecca Wittenstein, Kacy Sellers, and Christopher B. Swanson. 2011. "Improving Student Learning by Supporting Quality Teaching: Key Issues, Effective Strategies." Bethesda, MD: Editorial Projects in Education Research Center.
- Hill, David, "Guest column: Teacher Effectiveness Initiative is right path." *The Commercial Appeal*, November 11, 2011.
- Hinkin, Timothy and Chester Schriesheim. 2004. "If You Don't Hear from Me, You Know You are Doing Fine." *Cornell Hotel and Restaurant Administration Quarterly* 45.4: 362-372.
- Hogarth, R. M. 1980. *Judgment and choice: The psychology of decision*. Chichester, UK: Wiley.
- Holmes, Malcolm, and David Shand. 1995. "Management reforms: Some practitioner perspectives on the last ten years." *Governance* 8.4: 551-78.



- Honig, Meredith I. 2009. "No Small Thing: School District Central Office Bureaucracies and the Implementation of New Small Autonomous Schools Initiatives." *American Educational Research Journal* 46.2: 387-422.
- Hopson, Rodney K. 2005. "Reinventing Evaluation." *Anthropology and Education Quarterly* 36.3: 289-295.
- House, R., and J. Rizzo. 1972. "Role conflict and ambiguity as critical variables in a model of organizational behavior." *Organizational Behavior and Human Performance* 1: 467-505.
- Howard, H. 2000. *Managing Quality*. Broadstairs, Kent, UK: Scitech Educational.
- "Tennessee Teacher Evaluation System Report: SCORE Offers 7 Recommendations For Improvement," *The Huffington Post*, June 12, 2012. [http://www.huffingtonpost.com/2012/06/12/tennessee-teacher-evaluat\\_n\\_1590447.html](http://www.huffingtonpost.com/2012/06/12/tennessee-teacher-evaluat_n_1590447.html) (April 25, 2013).
- Ilgen, D. R., and J.M. Feldman. 1983. "Performance appraisal: A process focus." Eds. B. M. Staw & L. L. Cummings. *Research in organizational behavior*, Vol. 5, 141-196. Greenwich, CT: JAI Press.
- Ingersoll, R. 1998. "The problem of out-of-field teaching." *Phi Delta Kappan* 79.10: 773-776.
- Ingersoll, R. 2000. "A different approach to solving the teacher shortage problem." *Policy Perspectives* 2.2: 6, 8.
- Ingersoll, R. 2001. "Teacher turnover and teacher shortages." *American Educational Research Journal* 38.3: 499-534.
- Ingersoll, R., N. Alsalam, P. Quinn and S. Bobbitt. 1997. *Teacher professionalization and teacher commitment: A multilevel analysis*. Washington, DC: National Center for Education Statistics.
- Jackson, S. E. and R.S. Schuler. 1985. "A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings." *Organizational Behavior and Human Performance* 36: 16-78.
- Joyce, Philip G. 1996. "Appraising Budget Appraisal: Can You Take Politics Out of Budgeting." *Public Budgeting and Finance* 16.4: 21-25.
- Joyce, Philip G. 1997. "Using Performance Measures for Budgeting: A New Beat, or Is It the Same Old Tune?" In "Using Performance Measurement to Improve Public and Nonprofit Programs." 1997. Ed. Kathryn E. Newcomer. *New Directions for Evaluation* 75: 45-61.

San Francisco, CA: Jossey-Bass.

Judge, Timothy A. and Gerald R. Ferris. 1993. "Social Context of Performance Evaluation Decisions." *The Academy of Management Journal* 36.1: 80-105.

Kanter, Rosabeth Moss. 1983. *The Changemasters: Innovation for Productivity in the American Corporation*. New York: Simon and Schuster.

Kay, E.H., H. Meyer, and R.P. French. 1965. "Effects of Threat in a Performance Appraisal Interview." *Journal of Applied Psychology* 49.

Kee, Jed and Roger Black. 1985. "Is Excellence in the Public Sector Possible?" *Public Productivity Review* 9.1: 25-34.

Keith, T.Z., Reimers, T.M., Fehrmann, P.G., Pottbaum, S.M., & Aubey, L.W. 1986. "Parental involvement, homework, and TV times: Direct and indirect effects on high school achievement." *Journal of Educational Psychology* 78: 373-380.

Kelley, Michael, "Plan for merger of Memphis, Shelby County schools complete," *The Commercial Appeal*, June 17, 2012.

Kelley, Michael, "Memphis City Schools: A district born amid segregation, yellow fever fades into history," *The Commercial Appeal*, June 28, 2013.

Kettl, Donald F. 1990. "The Perils and Prospects of Public Administration." *Public Administration Review* 50.4: 411-419.

Kiel, Daniel. Personal Interview by Landon Webber, May 14, 2014.

Klein, Joel, "The Failure of American Schools," *The Atlantic*, April 26, 2011.

Kopczynski, M. and M. Lombardo. 1999. "Comparative Performance Measurement: Insights and Lessons Learned from a Consortium Effort." *Public Administration Review* 59.2: 124-134.

Koppich, Julia E. 2004. "Developing State Policy to Ensure a 'Highly Qualified' Teacher in Every Classroom." *NGA Center for Best Practices*.

Kouzmin, Alexander, Elke Löffler, Helmut Klages, and Nada Korac-Kakabadse. 1999. "Benchmarking and Performance Measurement in Public Sectors." *International Journal of Public Sector Management* 12.2: 121-144.

Kravchuk, R. S., and Schack, R. W. 1996. "Designing effective performance-measurement

- systems under the Government Performance and Results Act of 1993." *Public Administration Review* 348-358.
- Kuppermintz, Haggai. 2003. "Teacher Effects and Teacher Effectiveness: A Validity Investigation of the Tennessee Value Added Assessment System." *Educational Evaluation and Policy Analysis* 25:3: 287-298.
- Labovtitz, G.H. 1972. "More on Subjective Executive Appraisals: An Empirical Study." *Academy of Management Journal*.
- Ladd, Helen F. 2008. "Teacher effects: What do we know?" Eds. G. Duncan and J. Spillane. *Teacher quality: broadening and deepening the debate*, 3-28. Evanston, IL: Northwestern University Press.
- de Lancer Julnes, Patria and Marc Holzer. 2001. "Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation." *Public Administration Review* 61.6: 693-708.
- Landy, F.J. and J.L. Farr. 1980. "Performance Rating." *Psychological Bulletin* 87.1: 72-107.
- Landy, F. J. and J.L. Farr. 1983. *The Measurement of Work Performance: Methods, Theory and Applications*. Orlando, FL: Academic Press, Inc.
- Lapsley, I. 2009. "New Public Management: Cruellest Invention of the Human Spirit?" *Abacus* 1.45: 1-21.
- Leana, Carrie R. 2011. "The Missing Link in School Reform." *Stanford Social Innovation Review*, Fall.
- Leana, Carrie R. and Frits K. Pil. 2006. "Social Capital and Organizational Performance: Evidence from Urban Public Schools." *Organization Science* 17.3: 353-366.
- Lee, Cynthia. 1985. "Increasing Performance Appraisal Effectiveness: Matching Task Types, Appraisal Process, and Rater Training." *The Academy of Management Review* 10.2: 322-331.
- Lee, Valerie E., Anthony S. Bryk, and Julia B. Smith. 1993. "The Organization of Effective Secondary Schools." *Review of Research in Education* 19: 171-267.
- Leef, George, "A Key Reason Why American Students Do Poorly," *Forbes Magazine*, October 24, 2013.
- Leone, Robert A. 1984. Unpublished lecture notes, Kennedy School of Government, cited in Jed

- Kee and Roger Black's (1985) "Is Excellence in the Public Sector Possible?" *Public Productivity Review* 9.1: 25-34.
- Levin, Henry M. 1998. "Educational Performance Standards and the Economy." *Educational Researcher* 27.4: 4-10.
- Linn, Robert L. 2003. "Accountability: Responsibility and Reasonable Expectations." *Educational Researcher* 32.7: 3-13.
- Locher, A. H., and K.S. Teel. 1988. "Appraisal trends." *Personnel Journal* September: 139-145.
- Locke, E.A. and G.P. Latham. 1984. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.
- Locker, Richard, "Tennessee lawmakers approve teacher evaluation plan," *The Commercial Appeal*, January 15, 2010.
- Locker, Richard, "Proposed special district for Shelby County Schools faces legislative hurdles," *The Commercial Appeal*, November 28, 2010.
- Locker, Richard, "Tenn. Senate works on delay of school charter vote as Memphis lawmakers say hands off," *The Commercial Appeal*, January 27, 2011.
- Locker, Richard, "Teachers Call for Changes in Tennessee's New Evaluation System," *The Commercial Appeal*, January 18, 2012.
- Locker, Richard, "Haslam signs bill allowing local referendums on municipal schools in Shelby suburbs," *The Commercial Appeal*, May 9, 2012.
- Locker, Richard, "Gov. Haslam signs main municipal school district bill," *The Commercial Appeal*, April 24, 2013.
- Locker, Richard and Amos Maki, "Memphis Council's unanimous vote OKs surrender of charter by city schools," *The Commercial Appeal*, February 10, 2011.
- Longenecker, C. O. 1989. "Truth or Consequences: Politics and Performance Appraisals." *Business Horizons*, 1-7.
- Longenecker, C. O., D.A. Gioia and H.P. Sims. 1987. "Behind the Mask: The Politics of Employee Appraisal." *The Academy of Management Executive*, 1.3: 183-193.
- Longenecker, C. O., P.R. Liverpool, and K.A. Wilson. 1988. "An Assessment of Managerial/Subordinate Perceptions of Performance Appraisal Effectiveness." *Journal of Business*

- and Psychology*, 2.4: 311-320.
- Longenecker, Clinton, and Dean Ludwig. 1990. "Ethical Dilemmas in Performance Appraisal Revisited." *Journal of Business Ethics* 9.12: 961-969.
- Louis, Karen Seashore and Helen M. Marks. 1998. "Does Professional Community Affect the Classroom? Teachers' Work and Student Experiences in Restructuring Schools." *American Journal of Education* 106.4: 532-575.
- Lu, Adrienne. 2013. "How States Evaluate Teachers Varies Widely." *Stateline- The Daily News Service of The Pew Charitable Trusts*. <http://www.pewstates.org/project/stateline/headlines/how-states-evaluate-teachers-varies-widely-85899511032> (February 3, 2014).
- Mani, Bonnie. 2002. "Performance Appraisal Systems, Productivity, and Motivation: A Case Study." *Public Personnel Management* 31.2: 141-159.
- Marjoribanks, K. 1983. "The evaluation of a family learning model." *Studies in Educational Evaluation* 9: 343-351.
- Marlinga, Jennifer. 2006. "Continuous Feedback: An Analysis of Performance Evaluation and Employee Feedback Systems in a Trial Court Environment." Institute for Court Management.
- McGuinn, Patrick. 2012. "The State of Teacher Evaluation Reform: State Education Agency Capacity and the Implementation of New Teacher-Evaluation Systems." *Center for American Progress*.
- McKenna, Steve, Julia Richardson, and Laxmikant Manroop. 2011. "Alternative paradigms and the study and practice of performance management and evaluation." *Human Resource Management Review* 21.2: 148-157.
- McMillin, Zack, "Judge in merger suit rules Memphis City Schools will 'cease to exist' in 2013," *The Commercial Appeal*, August 8, 2011.
- McMillin, Zack, "Deal puts 23-member board in control of Memphis, Shelby County schools Oct. 1," *The Commercial Appeal*, August 24, 2011.
- McMillin, Zack, "Judge Mays' consent decree 'brings finality' to schools consolidation," *The Commercial Appeal*, September 29, 2011.
- McMillin, Zack, "How you define 'college ready' depends on the number," *The Commercial Appeal*, November 11, 2011.

- McMillin, Zack, "Shelby County municipal schools trial opens with revelation about loss of teacher rights," *The Commercial Appeal*, September 4, 2012.
- McMillin, Zack, "Judge Mays halts quest for municipal school districts," *The Commercial Appeal*, November 27, 2012.
- McMillin, Zack, "Shelby suburban leaders: No municipal schools for 2013-14," *The Commercial Appeal*, January 5, 2013.
- McMillin, Zack, "Shelby schools judge picks Rick Masson to oversee Memphis merger," *The Commercial Appeal*, March 5, 2013.
- McMillin, Zack and Jane Roberts, "Memphis voters OK school charter surrender," *The Commercial Appeal*, March 8, 2011.
- McMillin, Zack and Jane Roberts, "Shelby County, Memphis school boards unanimously OK merger deal," *The Commercial Appeal*, August 25, 2011.
- Meier, Kenneth. 2000. *Politics and the bureaucracy: Policymaking in the fourth branch of government*, 4th ed. Fort Worth, TX: Harcourt Publishing.
- Meyer, H. H., E. Kay, and J. French. 1965. "Split roles in performance appraisal." *Harvard Business Review* 43.1: 123-129.
- Memphis City Schools. 2007. "A Brief History of Memphis City Schools." *Memphis City Schools Website*. <http://96.4.164.27/Newsreleases/communications/MCS-History.html> (April 15, 2014).
- Memphis City Schools. 2010. "The Teacher Effectiveness Initiative." *Memphis City Schools Website*. [http://www.ourvoiceourschools.org/sites/346/uploaded/files\\_The\\_Teacher\\_Effectiveness\\_Initiative\\_Overview.pdf](http://www.ourvoiceourschools.org/sites/346/uploaded/files_The_Teacher_Effectiveness_Initiative_Overview.pdf) (April 25, 2013).
- Memphis City Schools. 2011. "Teacher Effectiveness Initiative: Top 5 Things You Need to Know About TEI and the Merger." *Memphis City Schools Website*. [http://www.mcsk12.net/tei/presskit\\_ntk.asp](http://www.mcsk12.net/tei/presskit_ntk.asp) (April 1, 2013).
- Memphis City Schools. 2012. "Employee Performance Evaluation Process." *Department of Human Resources, Office of Performance Management and Professional Development*.
- Memphis-Shelby County Education Association Staff (Anonymous Source). Personal Interview by Landon Webber, May 12, 2014.
- Midwinter, Arthur. 2008. "Performance Management and Best Value Audit in Scotland: A

- Research Note on Theory and Practice.” *Financial Accountability & Management* 24.4: 439-453.
- Millman, J., and H. D. Schalock. 1997. “Beginnings and introduction.” Ed. Jason Millman. *Grading teachers, grading schools: Is student achievement a valid evaluation measure*, 3-8. Newbury Park, CA: Corwin Publishing.
- Mohrman, Allan M., Susan M. Resnick-West and Edward E. Lawler III. 1989. *Designing Performance Appraisal Systems*. San Francisco: Jossey-Bass.
- Moynihan, Donald P. 2009. “Through A Glass, Darkly.” *Journal of Performance & Management Review* 32.4: 592-603.
- Moynihan, Donald P. and Sanjay K. Pandey. 2005. “Testing How Management Matters in an Era of Government by Performance Management.” *Journal of Public Administration Research and Theory: J-PART* 15.3: 421-439.
- Murphey, David A. 1999. “Presenting Community-Level Data in an ‘Outcomes and Indicators’ Framework: Lessons from Vermont's Experience.” *Public Administration Review* 59.1: 76-82.
- Murray, Michael A. 1975. ”Comparing Public and Private Management: An Exploratory Essay.” *Public Administration Review* 35: 364-371.
- Musolf, Lloyd D., and Harold Seidman. 1980. ”The Blurred Boundaries of Public Administration.” *Public Administration Review* 40: 124-130.
- Naisbett, John. 1982. *Megatrends*. New York: Warner Books.
- “Amid pushback, new teacher evaluation system has ally in Jesse Register,” *Nashville City Paper*, October 23, 2011.
- National Academy of Public Administration (NAPA). 1994. *Toward Useful Performance Measurement: Lessons Learned from Initial Pilot Performance Plans Prepared under the Government Performance and Results Act*. Washington, DC: NAPA.
- National Academy of Public Administration (NAPA). 1999. *Using Performance Data to Improve Government Effectiveness*. Washington, DC: NAPA.
- National Association for Alternative Certification. 2011. “NAAC Launches Retention Study.” (<http://www.alt-teachercert.org>), accessed 22 April 2014.
- National Association of State Boards of Education (NASBE). 1998. *The numbers game:*

- Ensuring quantity and quality in the teaching workforce*. Alexandria, VA: NASBE.
- National Center for Education Information. 2005. "Research Overview." (<http://www.ncei.com/overview.html>), accessed 22 April 2014.
- National Center for Education Statistics. 2007. "Parent and Family Involvement and Education." National Household Education Surveys Program. [http://nces.ed.gov/nhes/pdf/pfi/07\\_pfi.pdf](http://nces.ed.gov/nhes/pdf/pfi/07_pfi.pdf). (March 24, 2012).
- National Commission on Excellence in Education. 1983. "A Nation at Risk: The Imperative for Educational Reform."
- National Commission on Teaching and America's Future. 1996. *What matters most: Teaching for America's future*. New York: National Commission on Teaching and America's Future.
- National Education Association. 2010. "Research Spotlight on Alternative Routes To Teacher Certification." (<http://www.nea.org/tools/16578.htm>), accessed 22 April 2014.
- National Industrial Conference Board. 1964. "Personnel Practices in Factories and Office Manufacturing." *Studies in Personnel Policy* 194.
- National Institute for Excellence in Teaching. 2011. "TAP Teaching Standards Selected as Basis for Tennessee's New Statewide Teacher Evaluation Model." <http://www.tapsystem.org/newsroom/newsroom.taf?page=features&id=127> (April 25, 2013).
- National Institute for Excellence in Teaching. 2012. "TAP Research Summary." [http://www.tapsystem.org/publications/tap\\_research\\_summary\\_0210.pdf](http://www.tapsystem.org/publications/tap_research_summary_0210.pdf) (April 25, 2013).
- Nemko, Barbara and Harold Kwalwasser, "Why Teacher Colleges Get a Flunking Grade," *The Wall Street Journal*, 23 October 2013.
- New York City Department of Education. January 2013. "Division of Academics, Performance, & Support (DAPS) Performance Management Process-2012-13 School Year."
- Newmann, Fred M., M. Bruce King, and Peter Youngs. 2000. "Professional Development that Addresses School Capacity: Lessons from Urban Elementary Schools." *American Journal of Education* 108.4: 259-299.
- Newstrom, J.D. 1974. "Smooth the Way for Your Own Performance Appraisal." *Supervisory Management* 5.19.
- Niazi, A.A. 1976. "Performance Appraisal: An Approach." *Indian Management*.



- Nickols, Fred. 2000. "Don't Redesign Your Company's Performance Appraisal System, Scrap It! A Look at Costs and Benefits." *Corporate University Review*.
- Nye, B., Konstantopoulos, S., and L.V. Hedges. 2004. "How large are teacher effects?" *Educational Evaluation and Policy Analysis* 26.3: 237-257.
- O'Day, Jennifer A., and Marshall S. Smith. 1993. "Systemic reform and educational opportunity." Ed. Susan H. Fuhrman. *Designing coherent education policy: Improving the system*, 250-312. San Francisco, CA: Jossey-Bass.
- Osborne, David, and Peter Plastrik. 2000. *The Reinventor's Fieldbook: Tools for Transforming Your Government*. San Francisco, CA: Jossey-Bass.
- Padovani, Emanuele, Ana Yetano and Rebecca Levy Orelli. 2010. "Municipal Performance Measurement and Management in Practice: Which Factors Matter?" *Public Administration Quarterly* 34.4: 591-635.
- Pareek, U. 1976. "Interaction Exploration." Eds. Pfeiffer and Jones. *Annual Handbook for Group Facilitators*. University Associates of California.
- Performance Evaluation System Administrator (Anonymous Source). Personal Interview by Landon Webber and Kemmashela Smith, July 25, 2013.
- "Performance Management: HR Questions the Value of Appraisals." 2005. *Personnel Today*.
- Peter, T. J. and M. Austin. 1983. "A Passion for Excellence." *Fortune Magazine*, 20-33.
- Peters, Thomas J. and Robert H. Waterman, Jr. 1982. *In Search of Excellence: Lessons from America's Best Run Companies*. New York: Harper and Row.
- Pil, Frits K. and Carrie R. Leana. 2009. "Applying Organizational Research to Public Schools." *The Academy of Management Journal* 52.6: 1101-1124.
- Pohlmann, Marcus. 2008. *Opportunity Lost: Race and Poverty in the Memphis City Schools*. Knoxville, TN: University of Tennessee Press.
- Poister, Theodore H. 2010. "The Future of Strategic Planning in the Public Sector: Linking Strategic Management and Performance." *Public Administration Review Special Issue*: 246-254.
- Poister, Theodore H., and Gregory Streib. 1999. "Performance Measurement in Municipal Government: Assessing the State of the Practice." *Public Administration Review* 59.4: 325-35.

- Pollitt, C. and G. Bouckaert. 2004. *Public management reform: a comparative analysis*, 2nd ed. Oxford, UK: Oxford University Press.
- Pomerantz, E.M., Moorman, E.A., & Litwack, S.D. 2007. "The How, Whom, and Why of Parents' Involvement in Children's Academic Lives: More Is Not Always Better." *Review of Educational Research* 77: 373-410.
- Pritchard, R.D., P.O. Roth, S.D. Jones, P.J. Galgay, and M.D. Watson. 1988. "Designing a Goal-Setting System to Enhance Performance: A Practical Guide." *Organizational Dynamics* 17: 69-78.
- Rainey, Hal G., Sanjay K. Pandey, and Barry Bozeman. 1995. "Research note: Public and private managers' perceptions of red tape." *Public Administration Review* 55.6: 567-74.
- Rainey, H. G., and P. Steinbauer. 1999. "Galloping elephants: Developing elements of a theory of effective government organizations." *Journal of Public Administration Research and Theory*, 9: 1-32.
- Ravitch, D. 2010. *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Ravitch, D. 2012. "How to demoralize teachers." *Education Week* 23.
- Ravitch, Diane. 2012. "A Chance to Make History: What Works and What Doesn't in Providing an Excellent Education for All." *The New York Review of Books*, 17-19.
- Ravitch, Diane. 2012. "How, and how not, to improve the schools." *The New York Review of Books*.
- Ravitch, Diane. 2012. "No student left untested." *The New York Review of Books*.
- Ravitch, Diane and Joseph P. Viteritti. 2001. *Making Good Citizens: Education and Civil Society*. Binghamton, NY: Vail-Ballou Press.
- Rice, J. K. 2003. *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Rich, Motoko, "National Schools Debate Is on Display in Chicago," *The New York Times*, September 11, 2012.
- Rizzo, J., R. House, and S. Lirtzman. 1970. "Role conflict and ambiguity in complex organizations." *Administrative Science Quarterly* 15: 150-163.

- Roberts, Jane, "Memphis City Schools Accepts \$90 Million Gates Foundation Grant," *The Commercial Appeal*, November 18, 2009.
- Roberts, Jane, "Foes of Memphis, Shelby County school consolidation brace for fight," *The Commercial Appeal*, December 23, 2010.
- Roberts, Jane, "Bill in Works to 'Tweak' Teacher Evaluations in Tennessee," *The Commercial Appeal*, January 26, 2013.
- Roberts, Jane, "Memphis school consolidation will cost about 225 central office jobs," *The Commercial Appeal*, April 3, 2013.
- Roberts, Jane, "Central office staff getting pink slips week before merger final," *The Commercial Appeal*, June 22, 2013.
- Roberts, Jane, "School merger takes economic toll of major business closing up shop," *The Commercial Appeal*, June 25, 2013.
- Roberts, Katie Smith. 2011. "EOY Performance Management Review Training & Support Information." New York City Department of Education, Division of Academics, Performance, & Support.
- Rose, Kenneth H. 2005. *Project Quality Management: Why, What, and How*. Florida: J. Ross Publishing.
- Rosenholtz, Susan J. 1985. "Effective Schools: Interpreting the Evidence." *American Journal of Education* 93.3: 352-388.
- Roth, David and Watson Scott Swail. 2000. "Certification and Teacher Preparation in the United States." Pacific Resources for Education and Learning and the Educational Policy Institute.
- Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125.1: 175-214.
- Rothstein, Richard, et al. 2010. "Problems with the use of student test scores to evaluate teachers." *Economic Policy Institute*, 237-257.
- Sanders, William L. 2000. "Value-Added Assessment from Student Achievement Data: Opportunities and Hurdles." *Journal of Personnel Evaluation in Education* 14.4: 329-339.
- San Diego Unified School District, Parent Outreach & Engagement Department, Harold J.

- Ballard Parent Center. 2012. "Engagement vs. Involvement: What's The Difference?"
- Sayeed, Omer Bin, and P. V. Bhide. 2003. "Performance Appraisal Effectiveness: An Empirical Assessment." *Indian Journal of Industrial Relations* 39.2: 166-182.
- Schiflett, Kathy L. and Mary Zey. 1990. "Comparison of Characteristics of Private Product Producing Organizations and Public Service Organizations." *The Sociological Quarterly* 31.4: 569-583.
- Shelby County Board of Education. 2013. "Policy 4020: Performance Evaluation." *Shelby County Board of Education Policy Manual*. Effective July 1, 2013.
- Shelby County Schools. 2012. "Evaluation Form for Classified Employees." Shelby County Schools Human Resources.
- Shelby County Schools, Office of Performance Management Staff (Anonymous Source). Personal interview by Landon Webber, May 10, 2013.
- Shelby County Schools Teacher Focus Group Series (Anonymous Source). Individual Teacher Focus Group Responses, November, 2013.
- Shelby County Schools Teacher Focus Group Series (Anonymous Source). Individual Teacher Focus Group Responses, April, 2014.
- Sherman, A. W., G.W. Bohlander, and H.J. Chruden. 1988. *Managing Human Resources*. Cincinnati, OH: Southwestern Publishing.
- Shostack, G. Lynn. 1977. "Breaking Free From Product Marketing." *Journal of Marketing* 41 (1977): 73-80.
- Shrinkfield, A.J. and Stufflebeam, D.L. 1995. *Teacher evaluation: Guide to Effective Practice*. Evaluation in Education and Human Services. Norwell, MA: Kluwer Academic.
- Simon, Mark. 2012. "A tale of two districts." *Educational Leadership* 70.3: 58-63.
- Silence, Sherri Drake and Zack McMillin, "Shelby County leaders urge Memphis City Schools to revote; city board memberes say that's unlikely," *The Commercial Appeal*, January 3, 2011.
- Silence, Sherri Drake, "Shelby County school board chairman David Pickler wants boundaries set," *The Commercial Appeal*, November 30, 2010.
- Silence, Sherri Drake and Zack McMillin, "Shelby schools cleared for legal action to derail

- merger, but options nearing end of line,” *The Commercial Appeal*, January 6, 2011.
- Singh, P., A. Maggu and S. K. Warriar. 1981. “Performance Appraisal Systems: A Critical Analysis.” *Indian Journal of Industrial Relations* 16.3: 315-343.
- Sink, D. Scott, Sandra J. DeVries, and Thomas C. Tuttle. 1984. “Productivity Measurement and Evaluation: What is Available?” *National Productivity Review* 3.3: 265-287.
- Smith, Peter. 1990. “The Use of Performance Indicators in the Public Sector.” *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 153.1: 53-72.
- Smith, Marshall S. and Jennifer O’Day. 1991. “Systemic school reform.” Eds. S.H. Fuhrman and B. Malen. *The politics of curriculum and testing: The 1990 Yearbook of the Politics of Education Association*, 233-267. Bristol, PA: Palmer Press.
- Smith, Marshall S., Jennifer O’Day, and Susan H. Fuhrman. 1994. ”State Policy and Systemic School Reform 1.” *Systemic Change in Education*: 109.
- Smith, P.C. and L.M. Kendall. 1963. ”Retranslation of Expectation: An Approach to the Construction of Unambiguous Anchors for Rating Scales.” *Journal of Applied Psychology* 47.
- Smither, James W. and Manuel London. 2009. *Performance Management: Putting Research Into Action*. San Francisco: Jossey-Bass.
- Spiegel, W.R. and E.W. Mumma. 1961. "Merit-rating of Supervisors and Executives." Panel Study No. 14, Bureau of Business Research, University of Texas.
- Stevenson, D.L., & Baker, D.P. 1987. “The family-school relation and the child’s school performance.” *Special Issue: Schools and development. Child Development* 58: 1348-1357.
- Strauss, Jack. 2012. “The Economic Gains to Colorado of Amendment 66.” *Social Science Research Network*.
- Strauss, Valerie, “The deafening silence on test cheating,” *The Answer Sheet: The Washington Post Education Blog*. [www.washingtonpost.com/blogs/answer-sheet/post/the-deafening-silence-on-test-cheating/2012/08/11/d4d565e2-e2fb-11e1-a25e-15067bb31849\\_blog.html](http://www.washingtonpost.com/blogs/answer-sheet/post/the-deafening-silence-on-test-cheating/2012/08/11/d4d565e2-e2fb-11e1-a25e-15067bb31849_blog.html) (July 15, 2013).
- Streib, Gregory D. and Theodore H. Poister. 2002. “The Use of Strategic Planning in Municipal Governments.” *The Municipal Year Book, 2002*, 18-25. Washington, D.C.: International City/County Management Association.

- Taft, R. 1971. "The Ability to Judge People." Eds. W.W. Ronan, et al. *Perspectives on the Measurement of Human Performance*. New York, NY: Appleton Century Crofts.
- Taylor, R.L. and R.A. Zawacki. 1978. "Collaborative Goal Setting in Performance Appraisal: A Field experiment." *Public Personnel Management* 7.3: 1262-1270.
- Taylor, Sully and Mila Lazarova. 2009. "Boundaryless careers, social capital, knowledge management: Implications for organizational performance." *Journal of Organizational Behavior* 30.1: 119-139.
- Taylor, Susan M., Kay B. Tracy, Monika K. Renard, J. Kline Harrison and Stephen J. Carroll. 1995. "Due Process in Performance Appraisal: A Quasi-Experiment in Procedural Justice." *Administrative Science Quarterly* 40.3: 495-523.
- Tennessee Department of Education. 2010. "Tennessee Wins Race to the Top Grant." <http://news.tn.gov/node/4828> (April 25, 2013).
- Tennessee Department of Education. 2011. "About Tennessee First to the Top." <http://www.tn.gov/firsttothetop/about.html> (April 25, 2013).
- Tennessee Department of Education. 2011. "TVAAS: Tennessee Value-Added Assessment System." [http://www.tn.gov/education/assessment/doc/TVAAS\\_Fact\\_Sheet.pdf](http://www.tn.gov/education/assessment/doc/TVAAS_Fact_Sheet.pdf) (April 25, 2013).
- Tennessee Department of Education. 2012. "Teacher Evaluation in Tennessee: A Report on Year 1 Implementation." [http://www.tn.gov/education/doc/yr\\_1\\_tchr\\_eval\\_rpt.pdf](http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf) (April 25, 2013).
- Tennessee Department of Education. 2013. "Report Card and Archives." [http://www.tn.gov/education/data/report\\_card/index.shtml](http://www.tn.gov/education/data/report_card/index.shtml) (February 13, 2014).
- Theurer, J. 1998. "Seven pitfalls to avoid when establishing performance measures." *Public Management*. 80.7: 21.
- Toffler, Alan. 1980. *The Third Wave*. New York: Bantam Books.
- Torff, Bruce. 2005. "Getting It Wrong on Threats to Teacher Quality." *The Phi Delta Kappan* 87.4: 302-305.
- Ujifusa, Andrew, "Tennessee, Arne Duncan's Show Horse, Set to Delay PARCC Tests for One Year," *Education Week*, 2014.
- U.S. Department of Education. 2000. *Eliminating barriers to improving teaching*. Washington,

DC: U.S. Department of Education.

- Van Ryzin, Gregg G., and Etienne Charbonneau. 2010. "Public service use and perceived performance: An empirical note on the nature of the relationship." *Public Administration* 88.2: 551-563.
- Vevea, Becky, "New teacher ratings not dramatically different," *WBEZ Online*, September 18, 2013. <http://www.wbez.org/news/new-teacher-ratings-not-dramatically-different-108703> (January 15, 2014).
- Voke, Heather. 2003. "Chapter 1: Responding to the Teacher Shortage." *Keeping Good Teachers*. Ed. Marge Scherer. Alexandria, VA: Association for Supervision and Curriculum Development.
- Walker, Richard M. and George A. Boyne. 2006. "Public Management Reform and Organizational Performance: An Empirical Assessment of the U.K. Labour Government's Public Service Improvement Strategy." *Journal of Policy Analysis and Management* 25.2: 371-393.
- Wayne, A. J. 2000. "Teacher supply and demand: Surprises from primary research." *Education Policy Analysis Archives* 8.47.
- Weiss, Gail. 2004. "How to Give and Receive Employee Feedback: Tell Staffers Early and Often How They're Performing, and Listen to Their Job-Related Concerns." *Medical Economics* 81.10 : 78-80.
- Weiss, Joanne, "Education's 'Race to the Top' Begins," *Education Week*, 2010.
- Wholey, Joseph S., and Harry P. Hatry. 1992. "The Case for Performance Monitoring." *Public Administration Review* 52.6: 604-610.
- Wholey, Joseph S., and Kathryn E. Newcomer. 1997. "Clarifying Goals, Reporting Results." In "Using Performance Measurement to Improve Public and Nonprofit Programs." *New Directions for Evaluation* 75: 91-98. Ed. by Kathryn E. Newcomer. San Francisco, CA: Jossey-Bass.
- Wildavsky, Aaron. 1964. *The Politics of the Budgetary Process*. Boston: Little, Brown.
- Williams, R. et al. 1977. "International Review of Staff Appraisal Practices: Current Trends and Issues." *Public Personnel Management* 1.6.
- Wilson, James Q. 1989. *Bureaucracy: What government agencies do and why they do it*. New York, NY: Basic Books.

- Winters, Marcus A, "How to Grade Teachers," *National Affairs* 13, 2012.
- Wright, Bradley E., and Sanjay K. Pandey. 2011. "Public Organizations and Mission Valence When Does Mission Matter?" *Administration & Society* 43.1: 22-44.
- Yang, Kaifeng. 2009. "Examining Perceived Honest Performance Reporting by Public Organizations: Bureaucratic Politics and Organizational Practice." *Journal of Public Administration Research and Theory: J-PART* 19.1: 81-105.
- Yi, Karen, "Most teachers fare well in year two of new evaluation system," *Sun Sentinel*, December 3, 2013.
- Zedlewski, Edwin W. 1979. "Performance Measurement in Public Agencies: The Law Enforcement Evolution." *Public Administration Review* 39: 488-493.
- Zelinski, Andrea, "Tweaks to State Teacher Evals Planned," *Tennessee Report News Service, Inc.*, June 20, 2012.